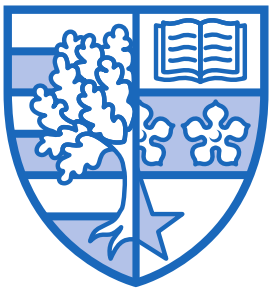
Optimal Design
of
Hybrid Optical Digital Imaging Systems

Tom Vettenburg

Submitted for the degree of Doctor of Philosophy



HERIOT-WATT UNIVERSITY

School of Engineering and Physical Sciences

December 2010

Abstract

Several types of pupil modulation have been reported to decrease the aberration variance of the modulation-transfer-function (MTF) in aberration-tolerant hybrid optical-digital imaging systems. It is common to enforce restorability constraints on the MTF, requiring trade of aberration-tolerance and noise-gain. In this thesis, instead of optimising specific MTF characteristics, the expected imaging-error of the joint design is minimised directly. This method is used to compare commonly used phase-modulation functions. The analysis shows how optimal imaging performance is obtained using moderate phase-modulation, and more importantly, it shows the relative merits of different functions. It is shown that the technique is readily integrable with off-the-shelf optical design software, which is demonstrated with the optimisation of a wide-angle reflective system with significant off-axis aberrations. The imaging error can also be minimised for amplitude-only masks. It is shown that phase aberrations in an imaging system can be mitigated using binary amplitude masks. This offers a low-cost, transmission-mode alternative to phase correction as used in active and adaptive optics. More efficient masks can be obtained by the optimisation of the imaging fidelity.

to Cristina,

Acknowledgments

This thesis would not have been conceivable without the help of many of the people I was fortunate to meet in Edinburgh. First of all I would like to thank my supervisor professor Andy R. Harvey for the opportunity to work on a project that combined my interests and experience perfectly. Thank you for excellent planning, and the many ideas and suggestions along the way that kept me focused. You made working on this thesis truly enjoyable for me.

The support of Qioptiq, St. Asaph has been pivotal for this project, both financially and scientifically. In particular I would like to thank Andy Wood and Nicholas Bustin for sharing their optical design knowledge and for their constant feedback that assured that the academic work remained of practical importance. Their many critical questions and insightful remarks have been invaluable. I would also like to thank Ian Hasler for providing thermal images of the constructed hybrid imaging system, and John Roberts for his valuable feedback when testing the simulator software.

This thesis would not have been the same without my colleagues. In particular I would like to thank Gonzalo, Mads, Alistair, and Bertrand for the many insightful discussions as well as for lending a hand in laboratory. Above all I am grateful to all of my colleagues for going out of their way to make me feel at home, both at the university and in the city of Edinburgh. Especially Mads for being a great friend and travel companion, and Claudio and Patrick for being excellent company in the office.

I would like to thank my parents for encouraging me since childhood to do what I enjoy most, and to give me every possibility for doing so. I am grateful to them for the innumerable opportunities I have been given throughout my life.

Most of all, I am deeply indebted to my loving wife Cristina for accepting and supporting my decision to change careers and country. Words cannot do justice to the unlimited patience and generosity she showed every day.

ACADEMIC REGISTRY
Research Thesis Submission

Name*:*	***Tom Vettenburg***		
School/PGI:	Engineering and Physical Sciences		
Version: *(i.e. First, Resubmission, Final)*	***Final***	Degree Sought (Award **and** Subject area)	***PhD Electrical***

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate*:*		Date:	

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in Academic Registry

Received in the Academic Registry by *(name in capitals):*			
Method of Submission *(Handed in to Academic Registry; posted through internal/external mail):*			
*E-thesis Submitted (**mandatory for final theses from January 2009**)*			
Signature:		Date:	

Please note this form should bound into the submitted thesis.

Updated February 2008, November 2008, February 2009

Contents

Publications

Journal Publications

Chapter 2 is adapted from the work published in the following paper:

"Fidelity optimisation for aberration-tolerant hybrid imaging systems"
T. Vettenburg, N. Bustin, and A. R. Harvey
26 Apr 2010, Optics Express, 18(9), pp. 9220-9228

Chapter 4 is adapted from two papers in preparation:

"Optimization and evaluation of radially-symmetric and antisymmetric phase-modulation for extending the depth-of-field of a hybrid imaging system"
T. Vettenburg, G. Muyo and A. R. Harvey
to be submitted to JOSA A.

"Holistic Optical-Digital Design – Wide-Field Reflective Imaging System"
T. Vettenburg and A. R. Harvey
to be submitted to JOSA A.

Chapter 5 is adapted from the work published in the following paper:

"Correction of optical phase aberrations using binary amplitude modulation"
T. Vettenburg and A. R. Harvey
25 Feb 2011, JOSA A, 28(3), pp. 429–433.

Conference Papers and Abstracts

"Achieving the potential of hybrid imaging systems"
A. R. Harvey, G. Muyo, M. Demenikov, and T. Vettenburg
27 Aug 2010, Proc. SPIE, Vol. 7818, 78180B (2010); doi:10.1117/12.868032

"Fidelity Comparison of Phase Masks for Hybrid Imaging"
T. Vettenburg, A. Wood, N. Bustin, and A. R. Harvey
15 Jun 2010, International Optical Design Conference, Jackson Hole, Wyoming, U.S.A.

"The issues of artefacts and noise in hybrid imaging systems"
A. R. Harvey, M. Demenikov, G. D. Muyo, T. Vettenburg, N. Bustin, I. Hasler, and A. Wood
7 Jun 2010, Imaging Systems, OSA technical Digest (OSA, 2010), paper IMD1.

"Optimisation and Application of Hybrid Optical-Digital Imaging Systems" *INVITED*
A. R. Harvey, M. Demenikov, G. D. Muyo, and T. Vettenburg
14 Oct 2009, Photonex09 Advanced Imaging and Microscopy, Coventry, U.K.

"Hybrid design of optical-digital imaging systems"
T. Vettenburg, M. Demenikov, G. Muyo, and A. R. Harvey
14 Oct 2009, Photonex09 Advanced Imaging and Microscopy, Coventry, U.K.

"Optimisation and Application of Hybrid Optical-Digital Imaging Systems" *INVITED*
A. R. Harvey, M. Demenikov, G. D. Muyo, and T. Vettenburg
11 Oct 2009, Frontiers in Optics 2009/Laser Science XXV,
OSA Technical Digest (CD) (OSA 2009), paper FThX2. San Jose, California, U.S.A.

"Optimisation in computational imaging systems"
N. Bustin, A. Wood, and T. Vettenburg
17 Sep 2009, U.K. Optical Designers' Meeting. Edinburgh, U.K.

"Optimality of pupil-phase profiles for increasing the defocus tolerance of hybrid digital-optical imaging systems" *FIRST AUTHOR INVITED*
T. Vettenburg, A. Wood, N. Bustin, and A. R. Harvey
21 Aug 2009, Proc. SPIE, Vol. 7429, 742903 (2009); San Diego, California, U.S.A.
doi:10.1117/12.825119

"Digital image processing as an integral component of optical design" *INVITED*
A. R. Harvey, T. Vettenburg, M. Demenikov, et al.
11 Sep 2008, Proc. SPIE, Vol. 7061 (2008); San Diego, California, U.S.A.
doi:10.1117/12.798581

Technical Article

"Optimising hybrid optical-digital imaging systems" *FIRST AUTHOR INVITED*
T. Vettenburg, M. Demenikov, G. Muyo and A. R. Harvey
16 Sep 2009, SPIE Newsroom. doi: 10.1117/2.1200909.1787

Internal deliverables

This research also led to the development of a comprehensive software library and toolkit for the rigourous simulation of hybrid optical-digital imaging systems. This software not only aided in the analysis leading to many of the results reported in this thesis, it has also been used to optimise an extended depth-of-field thermal imaging system developed at Qioptiq, St. Asaph. The optimised design has been fabricated and compared experimentally with the equivalent traditionally-designed thermal imaging system (cf. Chapter 4). The software is currently used at Qioptiq for the further analysis of novel hybrid imaging systems. A document titled "Manual to the Hybrid Optical-Digital Imaging System Simulation Library" has been provided to accompany the latest version of the software, to support the use of the software toolkit, and to facilitate its further development.

Chapter 1 – Introduction

1.1 Hybrid optical-digital imaging

The main objective of this thesis is to enable hybrid optical-digital imaging systems to achieve maximum aberration tolerance and image quality. This section explains what *hybrid* optical-digital imaging systems are, and how such systems can outperform conventionally-designed systems.

1.1.1 The digitisation of optical imaging systems

The history of imaging systems dates back to the *camera obscura* described by Aristotle, later constructed and studied by Alhazen [1]. Over the centuries lens and mirror-designs have been extensively developed along with manufacturing technology. A relatively recent evolution is the digitisation of the camera which has made imaging systems ubiquitous and more versatile than ever before.

The digitisation of imaging systems was sparked by the invention of electronic sensor arrays. The MOS[1] sensors [2] and CCD sensors[2] [3, 4] were first built at the end of the sixties. Although in principle a high contrast image can be created with the current technology; in practice, optical aberrations are often unavoidable. The object is not always located precisely at the focal plane, and constraints on size and cost hamper the reduction of the lens aberrations. Furthermore, aberrations can also be introduced by the media between the imaging system and the object, such as is the case for example with earth-based astronomy [5–7], imaging in the eye [8–12], or imaging through an aberrated window requiring conformal optics [13–15]. Few are the situations where precise control can be exerted over the object and environment, while lens dimensions and cost are irrelevant.

Fortunately, the tolerance to aberrations can be increased by transferring more of the burden to the digital electronics. Such *hybrid* optical-digital designs exploit the exponential growth in computational capacity seen in the last decades [16].

1.1.2 Hybrid optical-digital imaging systems

In 1972 Häusler proposed a technique to extend the depth-of-field of a microscope whereby the object is moved through the focal plane during the acquisition of a single

[1]The acronym (C)MOS stands for (Complementary) Metal-Oxide-Semiconductor and refer to the semiconductor technologies often used for the production of image sensor arrays found in consumer electronics.

[2]The Charge-Coupled Device, or CCD was invented in 1969 at AT&T Bell Labs by Willard Boyle and George E. Smith. It was not until 1975 that it was used by Steven Sasson at Kodak to create the first prototype of a digital camera.

image [17]. Häusler realised that although the through-focus-integrated image was blurry, a sharp and extended-depth-of-field image can be obtained by deconvolving it with the correspondingly-integrated point-spread function (PSF). More in general this showed that a sharp optical image is not strictly necessary for digital imaging systems, it is sufficient to record the information required for the digital restoration of a sharp image.

Dowski and Cathey showed that by placing a cubic phase mask at the aperture stop of an incoherent imaging system, this defocus-tolerant information can also be captured without moving the object [18]. The technique named *wavefront coding*™ has inspired the development of many phase and amplitude modulations for the mitigation of aberrations. Also the terms *pupil-phase-engineering* [19], and *point-spread function engineering* [20], have been used to refer to this concept. As these terms are sometimes used to refer to the application of a specific modulation type, the terminology *hybrid optical-digital imaging*, or *hybrid imaging* is preferred here to refer to any optical-digital imaging technique that does not rely on a sharp intermediate optical image.

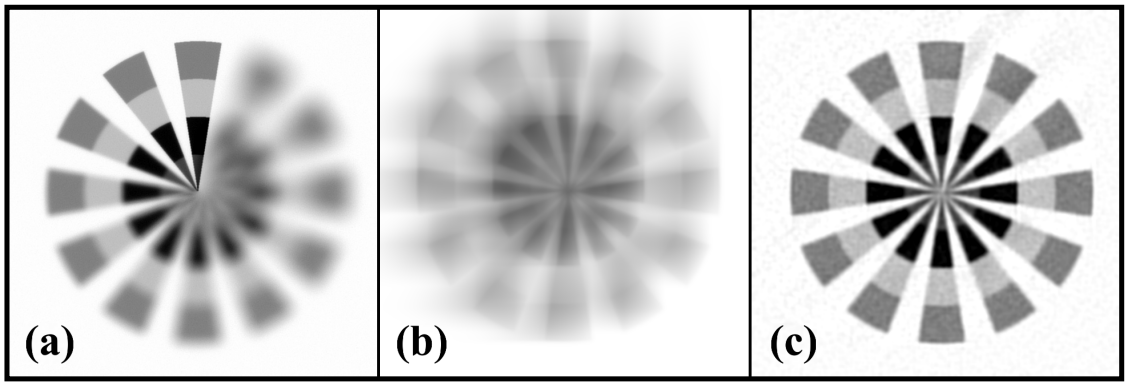


Figure 1.1: *Twelve spokes with anti-clockwise increasing defocus:* $0 \leq W_{20} \leq 4\lambda$*, imaged with various digital imaging systems with circular pupil: (a) a conventional diffraction-limited system, (b) a system with a cubic modulated pupil phase, (c) the same pupil-modulated imaging system applying image restoration. The intensity is inverted for best visibility; hence white indicates no irradiance, and black indicates detector saturation. A noisy detection is simulated by adding Gaussian noise with a standard deviation of 1% of the dynamic range of the image.*

As an example of the benefit of a hybrid design, figure 1.1 shows a synthetic depth-variant spoke target for a conventional imaging system (a), for a defocus-tolerant hybrid imaging system before image restoration (b), and after image restoration (c). The twelve o'clock spoke is in focus, and the amount of defocus of the spokes increases anti-clockwise to $W_{20} = 4\lambda$ at one o'clock. It can be seen from figure 1.1a that the spokes appear more and more blurred when the defocus increases, corresponding to what is expected for a conventional imaging systems. In contrast, cubic modulation of the pupil phase, discussed in chapter 2, causes all spokes in figure 1.1b to appear equally blurred.

The optical image shown in figure 1.1b is not immediately useful; however, as can be seen from figure 1.1c, acceptable contrast is obtained for all spokes simultaneously when applying image restoration. Note that the same was not possible for the image of the non-modulated system shown in figure 1.1a; here the equivalent image restoration only leads to a suppression of the high-frequency noise, and not to a contrast enhancement of the out-of-focus spokes.

Notice that the spokes in figure 1.1c are noisier and less sharp than the single in-focus spoke seen at the top of figure 1.1a. This is part of the trade-off made to obtain defocus-tolerance. A noisy detection process is simulated by adding Gaussian noise with a standard deviation of 1% of the dynamic range of the image. Significant noise amplification can occur at certain spatial frequencies; however, an appropriately regularised[3] image restoration algorithm will prevent that noise dominates the restored image. For a better understanding it is helpful to study the impulse response of the imaging system in the spatial-frequency domain.

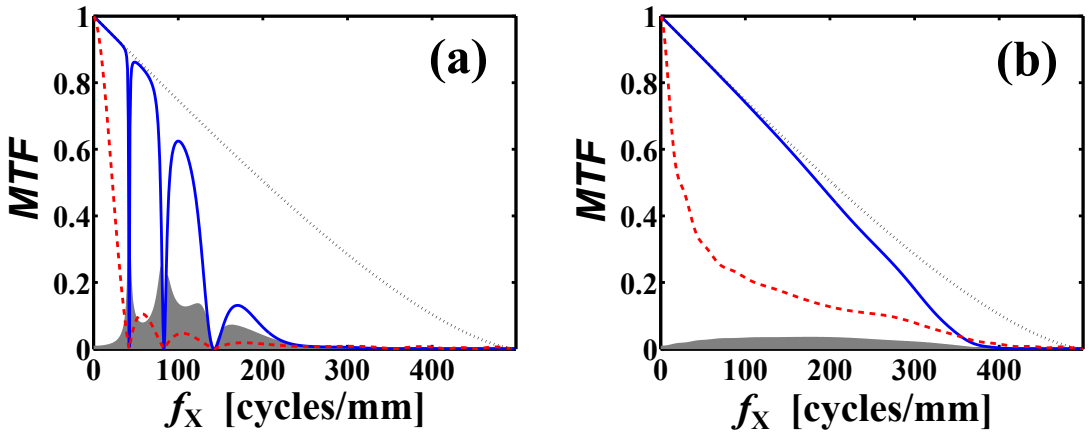


Figure 1.2: *The modulation transfer function (MTF) of two imaging systems, (a) without pupil-phase modulation, (b) with cubic phase-modulation. The dashed red line indicates the MTF before digital image restoration at a defocus of $W_{20} = 2\lambda$. The solid blue line shows the digitally restored MTF, and the magnitude after image restoration of 1% Gaussian noise is shown as a grey region at the plot abscissa. The diffraction-limited MTF is indicated by a dotted black line.*

As can be seen from the dashed red line in figure 1.2a, the modulation transfer function[4] (MTF) of a conventional system exhibits zeros when out-of-focus by two waves: $W_{20} = 2\lambda$. At these spatial frequencies no information is recorded, making

[3]Image restoration is typically an ill-posed problem. To guarantee a solution that is representative, even if the recording is noisy, it is necessary to include prior information so that the problem becomes well-posed. This is generally referred to as the *regularisation* of the ill-posed problem. Typically prior information about the signal-to-noise ratio is sufficient to obtain acceptable solutions.

[4]The modulation transfer function is defined as the absolute value of the optical-transfer function, which in turn is defined as the Fourier transform of the point-spread function normalised in intensity.

restoration without additional assumptions impossible[5]. The solid blue line indicates the MTF after restoration. Even when the optical MTF is non-zero, it could not be restored to the diffraction-limited level. A well-regularised image restoration algorithm will ensure that the correction it makes to the image is not cancelled by the degradation due to noise amplification. Image reconstruction is thus only possible at spatial frequencies where the MTF is sufficiently large to guarantee a high signal-to-noise ratio.

In contrast, the MTF of an out-of-focus pupil-modulated optical system is free of nulls, as can be seen from the dashed red line in figure 1.2b. Although the MTF is considerably below the diffraction-limited level (dotted black line), it remains sufficiently high to permit restoration (solid blue curve) to values near the diffraction-limited level. The general rule still holds that high MTF values are beneficial for image quality; however, for hybrid designs the correspondence is less direct and depends on the image restoration. Traditionally used metrics such as Strehl intensity ratio[6] lose therefore part of their significance.

1.1.3 The design of hybrid imaging systems

A digital-optical system can be regarded as an information channel. Information about the object is conveyed by photons scattered, transmitted, or radiated by the object, and through the optical path towards the sensor array. The irradiance is sampled at the detector and converted to a digital signal which could be used for electronic display or applications such as machine vision, bar-code reading, or automated object identification [22–24]. The digital optical system can record only a selected subset of all the information about the object; because it is filtered by the entrance pupil of the optics, the projection on the image plane, and the conversion to an electronic signal. The information channel has thus a limited capacity, and furthermore it is unreliable since aberrations affect the projection on the image plane.

From this viewpoint, hybrid imaging systems can be considered to optically *encode* the information before projection, making the transmission reliable, so that the electronics can *decode* the information unequivocally for the designed task. Such encoding necessarily introduces redundancy, and will therefore reduce the capacity of the channel. It can be understood that increased aberration-tolerance is associated with a reduction in the signal-to-noise ratio. The highest contrast can only be obtained with a diffraction-limited imaging system, in absence of aberrations.

[5]Non-linear restoration algorithms assuming constraints on the solution, such as positivity and smoothness of the irradiance – could be able to restore information to likely values at these spatial frequencies. A discussion of such algorithms can be found in most text books on image restoration, e.g. [21].

[6]The Strehl intensity ratio, I_S, is defined as the ratio between the peak-intensity of the PSF and that of the PSF of a diffraction-limited system.

Two points need to be addressed when designing hybrid imaging systems. The optimal form of pupil modulation must be decided for a particular application, and the performance of the system should be quantifiable in a non-subjective manner. An objective metric alone is not necessarily sufficient because in practice only a subset of all possible pupil modulation can be investigated within a reasonable time. In the following section, the most important pupil modulations are discussed along with the metrics used in their derivation.

1.2 Pupil modulation

The PSF of most optical systems can be considered shift-invariant over a small area of the field-of-view. It is convenient to maintain this property also for hybrid designs so that image reconstruction can be done efficiently in the spatial-frequency domain. Hence, most hybrid optical-digital designs modulate the light where it passes the aperture stop[7]. The modulation techniques discussed in the following subsection block or darken parts of the aperture, hence modulate the amplitude of the light that passes the pupil. In subsection 1.2.2, most of the common types of phase-modulation are discussed. Although phase-modulation often requires more complex and expensive manufacturing techniques, it has the advantage of a high light efficiency compared to amplitude modulation.

1.2.1 Pupil amplitude modulation

Many forms of pupil amplitude modulation have been proposed for the mitigation of aberrations, e.g. an *annular aperture* can be considered a binary amplitude attenuating mask. Already in 1960 Welford showed that these increase the depth-of-field [28]. Although the resolution is unaffected by such a mask, a considerable reduction in MTF can be expected at the mid-range spatial frequencies. Furthermore, the optical throughput can be severely reduced when using an annular aperture. Mino and Okano reported on an alternative approach to extend the depth-of-field whereby the transparency of the aperture is reduced quadratically from the centre to the rim [29]. Such attenuations are often referred to as *apodisations* because they reduce the intensity outside the Airy disc[8]. Similar apodisations for annular apertures have also been shown to improve the tolerance to defocus as well as spherical aberration [30, 31]. Although the techniques

[7] A notable exception is the plenoptic or light-field camera [25–27]. Here a lenslet array is placed before the sensor so that each microlens only illuminates a small part of the detector. The captured information permits reconstructing an image focused at the desired distance. Unfortunately the resolution of the image is typically reduced by a factor larger than 4.

[8] The PSF of a circular aperture is also referred to as the Airy disc or Airy pattern. This is mathematically defined as $I_0 \left(\frac{2J_1(r)}{r} \right)^2$, with r the distance from the centre and I_0 the central intensity. The intensity forms concentric rings around a high intensity disc-shaped region.

described in this paragraph guarantee a large Strehl intensity ratio and thus do not strictly require image processing, digital contrast enhancement can improve the image quality by reversing this MTF suppression inherent to any pupil modulation technique.

Also *aperture synthesis* as used in radio astronomy can be considered to be a form of pupil amplitude modulation whereby the whole aperture area in between antennas is blocked. This is an interesting technique because in principle arbitrary aberrations can be corrected for. In radio aperture synthesis large antenna arrays correlate the combined measurements to create a higher resolution signal. The smallest angle measurable is inversely proportional to the largest distance between two measurement points. Combining various antennas into an array will effectively create one large antenna with a higher resolution than each of the individual antennas. To create this high-resolution image, it is important that the signals of all antennas can be made to interfere. Until recently this effectively limited this technique for large apertures to radio frequencies. A detailed description of this principle can be found in [32].

In principle, aperture synthesis is tolerant to arbitrary aberrations because the spatial-frequency components of the scene are measured directly. Although an aberration affects the phase of the components, with the exception of tip and tilt, it can be reversed once all components are measured [33, 34]. Furthermore, low order aberrations tend to dominate in the scene, this allows noisy measurements to be regularised.

The aperture synthesis principle can be extended to optical wavelengths, however with the additional complication that the optical electromagnetic signals cannot be digitised before they interfere[9]. Electronic sampling and analog to digital conversion can be done at rates higher than the radio frequencies($< 100\,\mathrm{GHz}$), however not for visible or even low energy infrared light consisting of electromagnetic waves of considerably higher frequencies($> 500\,\mathrm{THz}$). By consequence, the optical signals have to be guided to the interference point by analog means. This is possible using fiber optics [35], or by using reflective optics.

Instead of guiding the light waves from different apertures to the image plane, one could also place the image plane far behind the individual apertures so that the light waves interfere by diffraction. In principle this is the same as having one large aperture that is completely obscured except for the places where the sub-apertures are located. This technique is usually referred to as *aperture masking interferometry* [33, 36–40], not to be confused with *coded aperture* [41–44], a technique that does not rely on interference and is typically used for high energy radiation.

In practice, aperture masking interferometry does not allow all spatial-frequency components in a single image to be measured unambiguously. A single Fourier com-

[9]The quasi monochromatic optical signal could in principle be sampled using optical heterodyning; however, terra-hertz sampling rates would be necessary to measure optical signals with a spectral width of no more than a few nanometer.

ponent can only be measured by a pair of appropriately distanced pinholes, sufficiently small so that the transmitted wavefront is approximately uniform. Certainly, several components can be sampled non-redundantly in a single frame; however, pinholes forming a continuous annulus would be required to sample the highest spatial frequencies, and this would already prevent unambiguous sampling of lower spatial frequencies. Furthermore, the optical throughput of such a pinhole mask is necessarily low. Redundant arrays with apertures of a size appropriate for the expected aberration are therefore deemed to be more practical.

Amplitude modulation has the potential to mitigate aberrations and can often be achieved at low cost; however, inherently it has the disadvantage that it limits the light throughput. One could consider amplitude masks that amplify the light selectively instead of attenuating it. This could be done for selected wavelengths by stimulated emission of photons from laser pumped elements in the mask. The dynamic range improvement probably does not justify the additional complexity. Furthermore, this would not reduce the uncertainty due to photon shot noise, which with increasing detector sensitivity is often the dominant source of noise. It must be of no surprise that no examples have been found in the literature using this technique. In what follows a more practical method to preserve light throughput is discussed.

1.2.2 Phase-modulation

A more recent technique to increase the tolerance to aberrations is modulating the optical path length, i.e. changing the relative phase of the passing light waves. Although typically more difficult to achieve, this method does not suffer from the inherent energy loss associated with amplitude masks. A sufficiently high light throughput is an essential requirement to obtain a good signal-to-noise ratio; however, it is not the only requirement. Equally, if not more, important is a sufficiently high MTF value at all spatial frequencies and independently of the aberration.

Choosing the appropriate phase-modulation depends on the specific requirements and is crucial for a system with good characteristics. Different types of phase-modulation have been reported for their potential to make the system less sensitive to various types of aberrations. The following paragraphs give an overview of the most common phase-modulations found in the literature.

Radial-symmetric

The fabrication of rotational-symmetric phase-modulations typically requires less complex manufactured techniques, hence reducing the production costs. An additional ad-

vantage of such profiles is that the optical-transfer function[10] (OTF) of the imaging system remains real; phase errors in the OTF are often considered visually more disturbing than equivalent changes in the MTF. The effect of the modulation is therefore only a contrast reduction which is relatively benign in terms of the perceived image quality [46, 47]. This type of phase-modulation might therefore even be usable without the need for image restoration, as suggested by Zalevsky et al. [48].

The *binary phase-modulation* proposed by Zalevsky et al. consists of three concentric parts as shown in figure 1.3a; an annulus causes a phase delay of a quarter wavelength compared to the central disc and the outer part of the aperture. This results in images with reduced sensitivity to defocus while maintaining a relatively strong central peak in the PSF. Since no digital post processing is required, such mask has recently been used in clinical trials for correcting presbyopia using contact lenses [49].

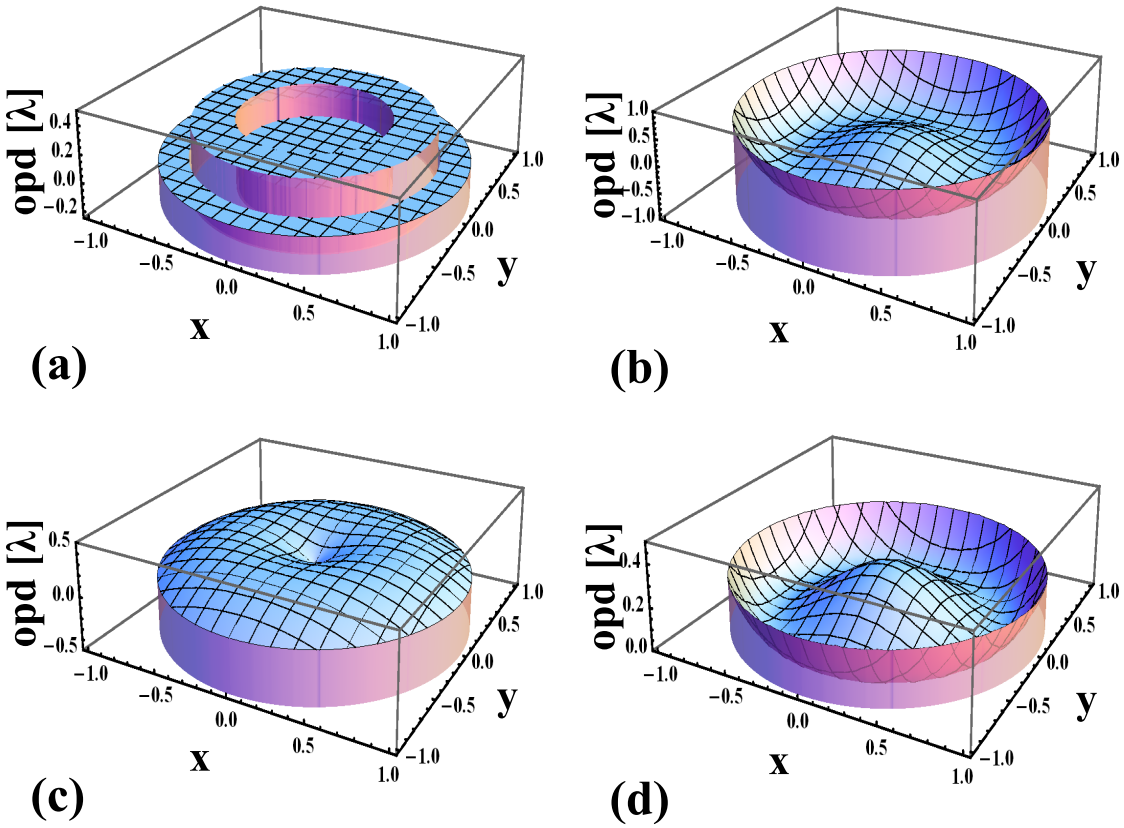


Figure 1.3: *Several radially-symmetric phase-modulations: (a) a binary phase-modulation as proposed by Zalevsky et al. [48], (b) the logarithmic asphere as derived by Chi and George [50], (c) the axicon [51], with a refocusing term added, and (d) the quartic phase-modulation [52].*

[10]The optical-transfer function is defined as the Fourier transform of the point-spread function, the image of a point source [45].

McLeod proposed the *axicon* which introduces a conical wavefront aberration to spread the intensity of a point source over a region between the lens and a point on the optical axis [51]. In principle such a phase-modulation can also be used for extended-depth-of-field imaging; however, the *logarithmic axicon* proposed by Sochacki et al. is often more practical since it focuses the light to a region between two points on the optical axis [53, 50, 54]. With a similar purpose, the *logarithmic asphere* shown in figure 1.3b was derived using Fermat's principle to image a line segment in object space to the same focal point [50, 54]. In principle also the focal region of the axicon can be adapted by adding an additional refocussing term as shown in figure 1.3c.

It is often more convenient to use a polynomial description of the radial phase-modulation. Fractional orders have been evaluated using simulations [55], and found to be optimal near the most commonly investigated fourth-order radial polynomial mask [56, 57], often referred to as the *quartic* [52], or spherical coding mask [58, 59]. The quartic mask depicted in figure 1.3d has been shown to have a on-axis intensity distribution that closely resembles that of the logarithmic asphere [52]. Higher order modulations have shown their benefit to compensate spherical aberration [52], and the off-axis aberrations coma and astigmatism [60].

Antisymmetric

Using the stationary-phase approximation Dowski and Cathey were able to deduce that the OTF becomes invariant with defocus with the introduction of a *cubic* phase-modulation at the pupil [18], such as shown in figure 1.4a. The invariance is however only guaranteed for a one-dimensional pupil and an asymptotically large cubic phase-modulation. The result can be extended to rectangularly-separable phase-modulations of the form: $\theta(x,y) = \alpha(x^3 + y^3)$, where x and y are cartesian coordinates on a rectangular pupil. However, the derivations only apply for spatial frequencies aligned with the x or y-axis; in practice much lower MTF values are found at other spatial frequencies.

The success of the cubic phase mask has inspired a range of phase-modulations resembling the cubic profile. In this category fall the *exponential* [61] and the *generic polynomial* phase-modulations [62]. The exponential type is optimised by minimising the derivative of the OTF with defocus, while constraining the integral of the MTF. The generic polynomial phase-modulation minimises the difference of the MTF with a target MTF while limiting the gradient of the phase-modulation profile. Both phase-modulations are thus obtained by making a rather arbitrary choice in how the MTF is bound. A higher MTF will improve the contrast and the restored image quality, while an MTF reduction typically makes it more tolerant to aberrations. Although both phase-modulations claim better defocus tolerance than the cubic phase-modulation, the comparison is not entirely convincing. The cubic has not been optimised using the same

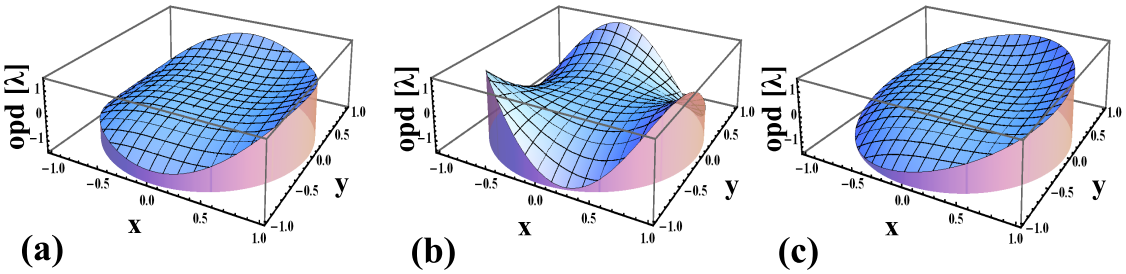


Figure 1.4: *The optical-path difference (OPD) introduced by phase-modulation of the form* $\theta(x,y) = \alpha(x^3+y^3)+\beta(x^2y+xy^2)$ *are shown for* $\alpha = 1\lambda$*; (a) the cubic phase-modulation with* $\beta = 0$*, (b) the generalised-cubic modulation* $\beta = -3\alpha$*, and, (c) the generalised-cubic modulation* $\beta = \alpha$*. Note that the normalised pupil coordinates u and v are typically several orders of magnitude larger than the wavelength used as unit for the optical-path-difference.*

metric and bounds on the MTF, and neither was the metric evaluated for the cubic profiles used in the comparisons.

Another similar profile is the *rectangularly separable, logarithmic* phase-modulation proposed by Sherif et al. [63], unrelated to the rotationally symmetric logarithmic phase-modulation discussed earlier. As it resembles the shape of the cubic modulation, it is not surprisingly that this phase-modulation also extends the depth-of-field. The separable logarithmic phase-modulation is claimed to outperform the cubic after optimisation with a metric that minimises the variation of the OTF with defocus and the difference with the diffraction-limited OTF at selected spatial frequencies. Differences are calculated using the Hilbert-space angle, and both terms are combined with a linear weighting factor. Although the MTF of the separable logarithmic phase-modulation is shown to have larger values at high spatial frequencies, it is unclear if this outweighs the reduction at low spatial frequencies.

By numerical optimisation of a two-dimensional third-degree polynomial phase-modulation Prasad et al. found that a phase-modulation of the form $\theta(x,y) = \alpha(x^3 + y^3)+\beta(x^2y+xy^2)$ would give better overall performance for $\beta \approx -3\alpha$ or $\beta \approx \alpha$ depending on the metric used [64]. The first expression is obtained by maximising the MTF, and minimising the PSF variance with defocus. The second expression is calculated my maximising the integral over spatial-frequency space of the minimum MTF value with defocus. Figures 1.4b and 1.4c show the respective phase profiles obtained by Prasad et al. Although the obtained masks appear to extend the depth-of-field for which good image quality can be obtained, it is unclear how to choose the weighting of the two terms in the first metric, while the metric used to derive the second profile is only based on the minimum MTF value with defocus, and ignores the phase of the OTF.

Other forms of phase-modulation

In an attempt to find the optimal phase-modulation computationally, a spline interpolation of eight points on a one-dimensional pupil phase has been optimised by Frauel and Castro using two different metrics[65]. The first metric is a minimisation of a linear combination of the maximum deviation of the OTF under defocus and the inverse of the in-focus MTF averaged over spatial frequencies. The obtained phase-modulation is similar to the cubic; however, somewhat lower in magnitude. By subjective analysis of the MTF, equal defocus invariance at higher contrast is claimed. The second metric was chosen to find a phase mask that can be used without image restoration, not surprisingly a phase-modulation of considerably lower magnitude is found; however, the defocus tolerance is lower as well.

Ashok and Neifeld found that pseudo random phase masks can increase the performance of an under-sampled system [66]. Their idea is inspired by the pseudo random code sequences used in code-division-multiple-access communication. The random, Gaussian-smoothed phase mask spreads the PSF over more pixels, which made it possible to restore the image with a resolution improvement of 50%[11], and a root-mean-square[12] (RMS) error reduction of 20%. Although random modulation is convenient to implement, it is unlikely that it is optimal for any given application.

1.2.3 Alternative PSF-engineering techniques

The simplicity of amplitude modulation, and the high light throughput of phase-modulation has made those the most commonly used forms of pupil modulation. Also *concurrent modulation of amplitude and phase* has been investigated [67–70]; however, light has two other properties that can be used in hybrid imaging: coherence length and polarisation.

Since annular apertures have a Strehl ratio that is less invariant with defocus, the incoherently combined image of various annular apertures must thus also be tolerant to defocus. Abrahamsson et al. showed that this can be achieved by placing concentric discs thicker than the coherence length at the aperture [71]. By definition the illumination of an incoherent imaging system has a short *coherence length*, hence this effectively splits the aperture in multiple apertures with complementary images that are added incoherently. As the image obtained with an annular aperture, the combined image is also less sensitive to defocus. It should be noted however that although the optical throughput is high, a

[11] Here the resolution is based on the Rayleigh criterion and calculated using simulations of a point object.

[12] The L^2 distance or root-mean-square difference between two functions $f(\mathbf{x})$ and $g(\mathbf{x})$ is defined as $\sqrt[2]{\int_\Omega |f(\mathbf{x}) - g(\mathbf{x})|^2 \mathrm{d}\mathbf{x}}$, where the integral is evaluated over the intersection, Ω, of the domains of $f(\mathbf{x})$ and $g(\mathbf{x})$.

sharp image can only be obtained if also the MTF has sufficiently high values at all spatial frequencies and defocus.

Similarly, two orthogonal *polarisation* states can be masked differently to improve some PSF characteristic. Chi et al. show that for unpolarised, or linearly polarised light[13], a focal depth extension can also be achieved by changing the polarisation in an outer annulus of the pupil [72]. The reasoning is that by changing the polarisation state of the light passing through one part of the aperture to the orthogonal polarisation, it will not interfere with the light passing the unaltered part. When a linear polariser would be placed before the image plane, it can be rotated to select either the light passing through the outer annulus or the light passing through the inner disc. Since for a defocus $W_{20} > \lambda/2$, the on-axis intensity can be increased either by making the aperture annular or by reducing its diameter; the same can be expected to be true when both images are combined in intensity as is the case for orthogonal polarisations; however, without discarding any light in the process.

The annular and Fresnel zone plate pattern were tested by Chi et al. [72] using a tunable spatial light modulator[14]. The annular mask extended the depth-of-field as predicted, without sacrificing light efficiency. The Fresnel zone plate is typically implemented using masks attenuating approximately 50% of the light, using the polarisation mask technique Chi et al. show that the light efficiency could effectively be doubled.

Circular and annular apertures have a depth-of-field inversely proportional to the open area [73, 28]. Apertures that partially rotate the light-polarisation, result in an image that is the same as the intensity sum of the two partial-aperture images. The obtained PSF and MTF are plotted in figures 1.5c and 1.5d respectively. Although the in-focus contrast is clearly reduced to that of the diffraction-limited system shown in figures 1.5a and 1.5b, the polarisation coded MTF at $W_{20} = 1/2\lambda$ (solid red line) and $W_{20} = 1\lambda$ (dashed green line) is free of nulls, permitting its digital restoration. Notice however that the defocus tolerance does not extend to $W_{20} = 2\lambda$, as can be seen from the nulls in the dotted blue line in 1.5d.

The polarisation coding is interesting because it can be seen, both as a binary amplitude modulation, or as a binary phase-modulation technique. Chi et al. studied it as a amplitude modulation technique; however, the depth-of-field extension can also be understood by considering that the polarisation modulating element delays one polarisation component by half a wavelength with respect to its orthogonal component when the

[13] More generally it can be seen that the degree of polarisation must be either 0 or 1, hence also circular polarisation can be used in the system proposed by Chi et al. The system however does not perform equally well under partially polarised light, which could also pose problems when polarising objects are present in unpolarised-illuminated scenes.

[14] The polarisation axis of light rotates as it traverses through a birefringent material. Liquid crystal displays use this property together with polarisation filters to selectively block light. By using these liquid crystals without the polarisers it functions as a tunable polarisation mask.

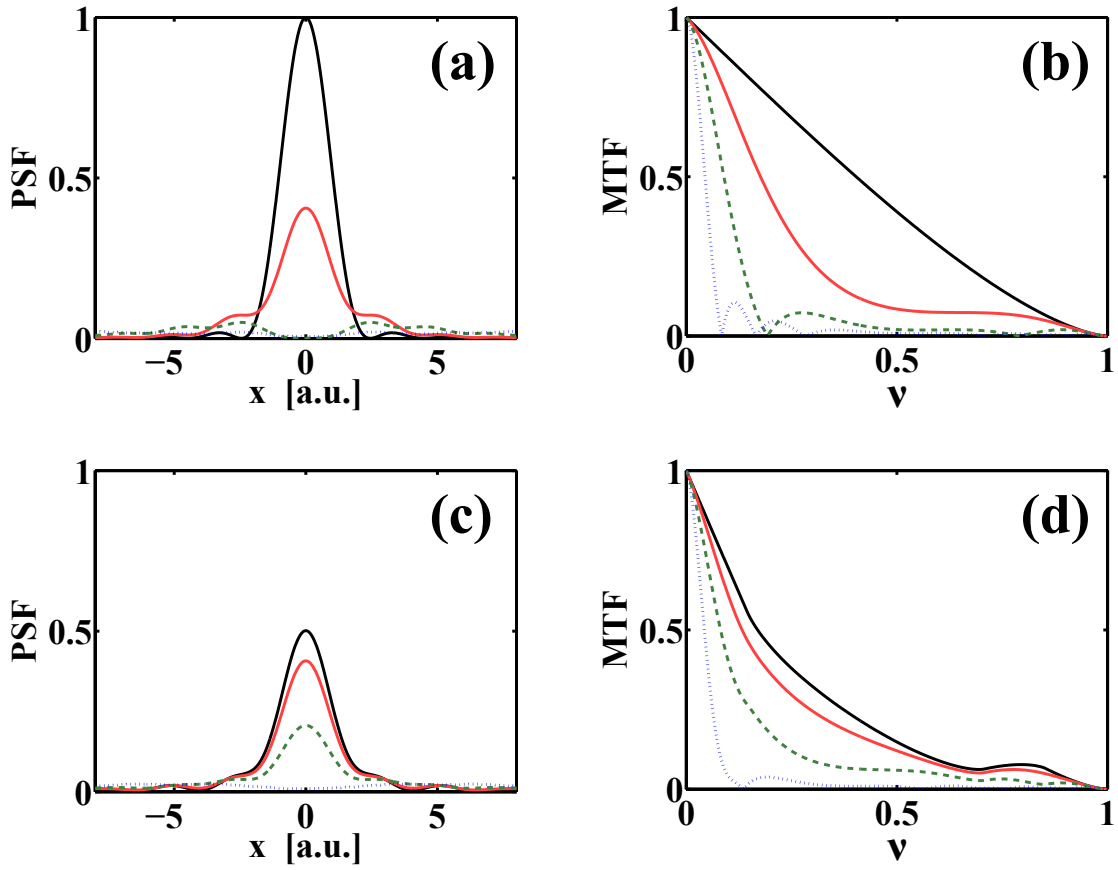


Figure 1.5: *The PSF (a,c) and MTF (b,d) are shown for a conventional aperture (a-b), and a polarisation-masked aperture (c-d), for various amounts of defocus: $W_{20} = 0$ (solid black line), $1/2\lambda$ (solid red line), 1λ (dashed green line), 2λ (dotted blue line). The PSFs are normalised so that the maximum irradiance of the non-aberrated system without mask is unity. The units of distance x are $\lambda/4\mathrm{NA}$, where* NA *is the numerical aperture; hence the PSF is Nyquist-sampled at integer values of x.*

appropriate basis is chosen. It can be seen that at least one of the polarisation components has a maximum phase difference of half a wave for a defocus as large as $W_{20} = 1\lambda$, so that every two points in the pupil interfere constructively. Destructive interference does appear for the coded pupil at $W_{20} > 1\lambda$, as it does for the non-coded system at $W_{20} > \lambda/2$.

1.3 Outline and achievements

Crucial to the design of a hybrid imaging system is the correct choice of pupil modulation. As seen in the previous section, a multitude of metrics has been used to design and compare different types of pupil modulation. Although these metrics can give an indication of the restored image quality, no direct relationship exists. It therefore remains unclear if, or under which conditions, the optimisation of these metrics would lead to the highest fidelity hybrid imaging system.

One of the most prominent metric is the *Strehl Intensity ratio*. Its success can be explained by the fact that it can be derived analytically for many systems of interest. Rayleigh found that if the maximum pupil deviation for spherical aberration is within a quarter wavelength then $I_S \geq 0.80$, it was later found this rule can also be applied to most other aberrations [74, p. 468]. Marechal made a similar link between I_S and the RMS wavefront error, which should be below approximately $\lambda/14$ to achieve $I_S \geq 0.8$.

While the Strehl ratio is a practical metric in traditional optical design, it has its limitations when applied to a hybrid imaging system. This is more apparent when it is viewed as the ratio of the integral of the aberrated optical-transfer function (OTF) and the diffraction-limited OTF. The post-processing step of hybrid imaging systems can cancel the phase error in the OTF, hence the phase should be insignificant to a hybrid imaging metric. However, unless the OTF phase change is proportional to the spatial-frequency, it will have a considerable effect on the Strehl ratio. Furthermore, a reduction to zero of the MTF at one spatial-frequency can be compensated by an equal increase at another spatial-frequency. While there is a clear loss of information at the first spatial-frequency, an equal gain at the other spatial-frequency is less valuable if the MTF is already sufficiently above the noise level.

Phase-modulations have been derived using various measures of the distance between PSFs or MTFs, e.g. the L^1 or L^2-norm, Fisher information[15], or the Hilbert-space angle. While this more readily translates to the digital restorability, it remains unclear if this provides any quantitative measure of the fidelity of the restored image. As appears in this thesis, more appropriate metrics exist.

A more direct measure of image quality can be the *root-mean-square* (RMS) *difference* between the image and the ideal diffraction-limited image. For linear systems this error will be proportional to the image irradiance, it is therefore more practical to normalise the measure against the irradiance of the image. For linear and shift-invariant imaging systems the RMS difference between the measured PSF and the diffraction-limited PSF can be used without having to assume anything about the imaged scene. Although the RMS error is straightforward to use, it is not necessarily a good indication of perceptual image difference. E.g. shifting the image by merely one pixel will result in a large RMS error, while the images still appear identical. Furthermore, the RMS error depends on the irradiance distribution of the object being imaged.

In chapter 2 it is shown how the expected RMS error can be calculated using a stochastic model of the scene. This permits for the first time a quantitative comparison of different types of pupil modulation based directly on image-fidelity. The proposed

[15] An optical system can be considered defocus insensitive when its PSF or OTF provide little information about the amount of defocus present. This can be quantified using Fisher information, defined as the expectation value of $\left(\partial \log P(W_{20})/\partial W_{20}\right)^2$, where $P(W_{20})$ is the probability of a certain observation value as a function of defocus, W_{20}.

method is applied to the generalised-cubic phase-modulations for which two local optima are found with a linear dependency between the maximum phase-modulation depth and the maximum defocus to-be-tolerated. To assess the relevance to perceived image quality, the mean structural similarity metric (MSSIM) has been used to assess image quality. This metric has been shown to correlate well with human perception of image similarity [75].

In chapter 3 it is shown how the model proposed in chapter 2 can be refined to account for more realistic simulations. In particular the integration with off-the-shelf optical-design software is discussed to facilitate the simulation of non-paraxial, thick lenses. Also the influence of broadband illumination and the polarised nature of electromagnetic radiation on pupil-modulated optical systems is discussed.

These principles are then applied in chapter 4, showing the practicality of its integration with the optical-design process. The first application presented in this chapter optimises and compares various commonly used phase-modulations for their ability to increase the tolerance to defocus. Results are shown for a paraxial Nyquist-sampled system, and for a LWIR system with four thick lenses exhibiting significant aliasing. As a second application, the field-of-view width of a two mirror reflective system is increased to 30° by optimising the hybrid imaging fidelity across the entire field-of-view volume.

In chapter 5, amplitude-only modulation is studied for its ability to mitigate aberrations. Although amplitude modulation has a reduced optical efficiency, it is often less expensive to implement than phase-modulation. In particular dynamic binary amplitude modulation can be achieved at relatively low cost for a broad range of wavelengths [76–79]. First, a theoretical upper limit to the aberration mitigation ability of an amplitude mask is calculated. In the limit for large aberrations, the MTF is derived for a specific type of mask, termed the *contour mask*, and shown to be sufficient to permit digital image restoration. Furthermore, it is shown that for finite aberrations the derived MTF can be surpassed by optimisation of the image-fidelity.

In the concluding chapter, our findings are summarised and the impact of this work on the design of hybrid optical-digital imaging systems is discussed.

Chapter 2 – Performance of idealised hybrid imaging systems

2.1 Introduction

Optical designs typically consist of many elements and can be analysed only in approximation with the help of analytical methods. Optical designers therefore rely on computer-aided design tools that allow numerical evaluation of the constructed model. Such software not only avoids the construction of prototypes early in the design process, it also permits automated optimisation and the prediction of the impact of manufacturing tolerances. This requires the availability of metrics that efficiently and accurately predict the performance of the optical design. The definition of performance is usually a combination of cost-related factors, and image quality requirements. In this chapter it will be shown how the imaging fidelity can be predicted efficiently and objectively for a hybrid imaging system. The method allows for the first time an objective comparison of hybrid imaging systems based directly on imaging fidelity. This is demonstrated with a quantitative comparison of three of the most common phase-modulations used to increase the tolerance to defocus.

As noted in the introduction, the metrics used in traditional optical design are less useful when designing hybrid optical-digital imaging systems. Traditional metrics such as spot-radius and wavefront error assume a direct relationship between aberrations and image quality. While this is a reasonable assumption for the optical image, the same is not necessarily true for the digitally processed image. The quality of a well-restored image depends only on the information carried in the recorded image, not on how this information is contained in the optical image. Certainly, aberrations have a significant influence on the amount of information passed to the image plane, however a direct relationship is not obvious. Furthermore, the information contained in the recorded image also depends on what scene is being imaged, and on the errors or noise introduced before, during, or after image acquisition.

A practical imaging system metric must be indicative of the quality of the images it produces. Note however that the definition of quality will necessarily depend on the purpose of the imaging system. For example, a visually pleasing image is not necessarily required if the system is only to be used to localise objects. An identical, but shifted image will be perceived of equal sharpness; however, it will be of less-value for object localisation purposes. Other examples are systems relying on automated image recognition which extract specific information from the image and hence the image recorded by such systems need not be perceived as sharp [80]. This observation has been exploited in [81, 82] to design iris recognition systems that require less accurate positioning of the subject's eye. More specifically, a task-specific metric involving a recognition algorithm was used to evaluate the performance of a hybrid optical-digital

system that combined pupil-phase modulation with object recognition. Following a similar approach, an extended-depth-of-field bar-code reader has been designed by Milgrom et al. [83]. It relies on chromatic defocus so that, independently of the object distance, the bar-code can be decoded from the information in the red, green, and blue channel of the sensor array. The depth-of-field extension is however only as large as the number of wavelength bands that the detector can discern and that can be guaranteed to be present in the illuminated bar-code.

In general however, most applications require that the image is close to the geometric projection of the scene onto the image plane. Specific applications merely differ in the relative importance given to the various possible deviations from this ideal geometric image. In this thesis, the important class of imaging systems is considered for which the final image must be perceived by a human observer as an accurate representation of reality. A well-defined, objective metric of image fidelity is however required to permit the automated design of imaging systems.

The meaning of imaging fidelity of a hybrid system is formalised in the following section, while in section 2.3 we discuss the different sources of noise and how these can be accounted for by an imaging fidelity metric. It is seen that the hybrid imaging fidelity not only depends on the OTF and the noise level; it also depends on the signal, i.e. the irradiance of the to-be-imaged object. Although the signal is highly application specific, in section 2.4 it is discussed how general characteristics of the scene can be sufficient to enable the estimation of imaging fidelity. In section 2.5 this is applied to evaluate and compare hybrid imaging systems incorporating various common pupil-phase modulations to mitigate defocus, which leads to some general design rules for this type of hybrid imaging systems. As will be shown in chapter 3, the presented method can be integrated into existing optical-design software enabling a holistic optical-digital design approach.

2.2 Evaluation of hybrid imaging fidelity

A hybrid imaging system can be considered to perform well if the final restored image is similar to the image of a diffraction-limited optical system. To quantify performance, the expected mean-squared error between the ideal and restored images is calculated over the required aberration tolerance range. Imaging fidelity is here defined as the inverse of the expected mean-squared error, $\bar{\varepsilon}$, between the restored image, $i_r(x,y)$, and the noiseless diffraction-limited image, $i_{\mathrm{dl}}(x,y)$:

$$\varepsilon^2 = \left\langle \mathbb{E}\left(\left| i_r(x,y) - i_{\mathrm{dl}}(x,y) \right|^2 \right) \right\rangle_{x,y} \tag{2.1}$$

$$\bar{\varepsilon} = \sqrt{\langle \varepsilon^2 \rangle_{\mathrm{ab}}}. \tag{2.2}$$

The symbol $\mathbb{E}(\cdot)$ denotes the expectation value, and the triangular brackets $\langle\cdot\rangle_{x,y}$ and $\langle\cdot\rangle_{\mathrm{ab}}$ indicate the ensemble averages over all pixels and the aberration tolerance range respectively. Using Parseval's equality, equation (2.1) can be rewritten in the spatial-frequency domain as:

$$\varepsilon^2 = \left\langle \mathbb{E}\left(|I_r(f_\mathrm{X}, f_\mathrm{Y}) - I_\mathrm{dl}(f_\mathrm{X}, f_\mathrm{Y})|^2 \right) \right\rangle_{f_\mathrm{X}, f_\mathrm{Y}}, \tag{2.3}$$

where the Fourier-transform of the restored image, $i_r(x,y)$, and the noiseless diffraction-limited image, $i_\mathrm{dl}(x,y)$, are represented by $I_r(f_\mathrm{X}, f_\mathrm{Y})$, and $I_\mathrm{dl}(f_\mathrm{X}, f_\mathrm{Y})$ respectively. The ensemble average is now calculated over the spatial frequencies f_X and f_Y defined by the detector array.

Assuming that the image formation process is linear and shift-invariant, the Fourier transform of the acquired image can be written as:

$$I_\mathrm{acq}(f_\mathrm{X}, f_\mathrm{Y}) = H_\mathrm{ab}(f_\mathrm{X}, f_\mathrm{Y}) S(f_\mathrm{X}, f_\mathrm{Y}) + N(f_\mathrm{X}, f_\mathrm{Y}), \tag{2.4}$$

where $H_\mathrm{ab}(f_\mathrm{X}, f_\mathrm{Y})$ is the optical-transfer function of the aberrated optics including any intensional pupil-phase modulation, and the functions $N(f_\mathrm{X}, f_\mathrm{Y})$ and $S(f_\mathrm{X}, f_\mathrm{Y})$ represent the Fourier transform of respectively the noise and the geometrical projection of the scene irradiance, both expressed in the equivalent grey levels. The mean-squared-error minimising filter, often referred to as the Wiener filter, $H_\mathrm{w}(f_\mathrm{X}, f_\mathrm{Y})$ minimises the expression $\left(H_\mathrm{w}(f_\mathrm{X}, f_\mathrm{Y}) I_\mathrm{acq}(f_\mathrm{X}, f_\mathrm{Y}) - I_\mathrm{dl}(f_\mathrm{X}, f_\mathrm{Y})\right)^2$ if the noise $N(f_\mathrm{X}, f_\mathrm{Y})$ has zero mean and is independent of $S(f_\mathrm{X}, f_\mathrm{Y})$. For readability omitting the function arguments $(f_\mathrm{X}, f_\mathrm{Y})$, the Wiener filter correcting for an aberration H_ab is given by:

$$H_\mathrm{w} = \frac{H_\mathrm{dl} H_\mathrm{ab}^*}{|H_\mathrm{ab}|^2 + {}^{P_N}/_{P_S}}, \tag{2.5}$$

where the symbol * indicates the complex conjugate, and the power spectral density of the scene is defined as $P_S = \mathbb{E}\left(|S|^2\right)$, and that of the noise as $P_N = \mathbb{E}\left(|N|^2\right)$. The imaging error can thus be minimised by setting $I_\mathrm{r} = H_\mathrm{w} I_\mathrm{acq}$ in equation (2.3):

$$\varepsilon^2 = \left\langle \mathbb{E}\left(\left| H_\mathrm{w} I_\mathrm{acq} - I_\mathrm{dl} \right|^2 \right) \right\rangle_{f_\mathrm{X}, f_\mathrm{Y}}, \tag{2.6}$$

which after substitution of I_acq given by equation (2.4), and the noiseless diffraction-limited image defined as $I_\mathrm{dl} = H_\mathrm{dl} S$, results in:

$$\varepsilon^2 = \left\langle \mathbb{E}\left(|H_\mathrm{w}\left(H_\mathrm{ab} S + N\right) - I_\mathrm{dl}|^2 \right) \right\rangle_{f_\mathrm{X}, f_\mathrm{Y}} \tag{2.7}$$

$$= \left\langle \mathbb{E}\left(|\left(H_\mathrm{w} H_\mathrm{ab} - H_\mathrm{dl}\right) S + H_\mathrm{w} N|^2 \right) \right\rangle_{f_\mathrm{X}, f_\mathrm{Y}}. \tag{2.8}$$

If, as already assumed for the Wiener filter, N has mean zero and is independent of

S, then the imaging error, ε, can be simplified to a sum of an error due to regularisation, and an error due to noise amplification as follows:

$$\varepsilon^2 = \left\langle \left| H_{\mathrm{w}} H_{\mathrm{ab}} - H_{\mathrm{dl}} \right|^2 P_S \right\rangle_{f_{\mathrm{X}}, f_{\mathrm{Y}}} + \left\langle \left| H_{\mathrm{w}} \right|^2 P_N \right\rangle_{f_{\mathrm{X}}, f_{\mathrm{Y}}} . \tag{2.9}$$

Notice, that equation 2.9 shows that exact knowledge of the scene and noise is not required to estimate the imaging error. As sections 2.3 and 2.4 will show, it is convenient that the argument of the Fourier components has no influence on the metric, because it enables an estimate for P_S and P_N, and by consequence also an estimate for ε.

Equation (2.9) shows that the imaging error is a combination of the restoration error weighted by the magnitude of the signal, and the filter weighted by the magnitude of the noise. Typical variations of $|H_{\mathrm{ab}}|$ and $|H_{\mathrm{w}}|$ with normalised spatial-frequency, are shown in the left-hand column of figure 2.1, for an optical system with a defocus of $W_{20} = 1\lambda$. As shown in figure 2.1(a), the MTF of an out-of-focus system decreases to zero at only a fraction of the cut-off spatial-frequency and it remains low for higher spatial frequencies. The corresponding Wiener filter shown in figure 2.1(c) amplifies the Fourier components where the signal-to-noise ratio is sufficiently large, resulting in the restored transfer function $H_{\mathrm{w}} H_{\mathrm{ab}}$ shown in absolute value in figure 2.1(e). The difference between the restored OTF and the diffraction-limited OTF is shown in absolute value in figure 2.1(g). It can be noticed that the transfer function is suppressed at high spatial frequencies due to the low signal-to-noise ratio, and around specific spatial frequencies where the MTF has nulls.

The right-hand column of figure 2.1 shows the same plots for an equally out-of-focus system incorporating a cubic phase-modulation of magnitude $\alpha = \lambda/2$, as given by equation 2.14 on page 30. It can be seen from figure 2.1(b) that, in particular along the axes ν_{X} and ν_{Y}, the MTF has appreciable values for a larger fraction of the spatial-frequency cut-off. A considerable part of the spatial-frequency spectrum can be restored by multiplication with the Wiener filter shown in figure 2.1(d), resulting in the restored transfer function shown in figure 2.1(f). Although the highest spatial frequencies could not be recovered due to the low signal-to-noise ratio, as can be appreciated from figure 2.1(h), a consistently low imaging error is obtained for the lower spatial frequencies. This is further discussed in section 2.5.3.

Since the Wiener filter H_{w} is a function of H_{ab} and the fraction P_S/P_N, knowledge of the variance of the scene spectrum, P_S, and noise spectrum, P_N is sufficient to predict the imaging error of a given system. In section 2.3 it is discussed why often the spatial-frequency spectrum of the noise is modelled as white and Gaussian. In this case the comparison of hybrid imaging systems can be based solely on H_{ab} and P_S/P_N. While the latter is scene-dependent, a representative measure of ε can be obtained by means of a typical amplitude spectrum of scenes. In section 2.4 it is discussed how such a spectrum

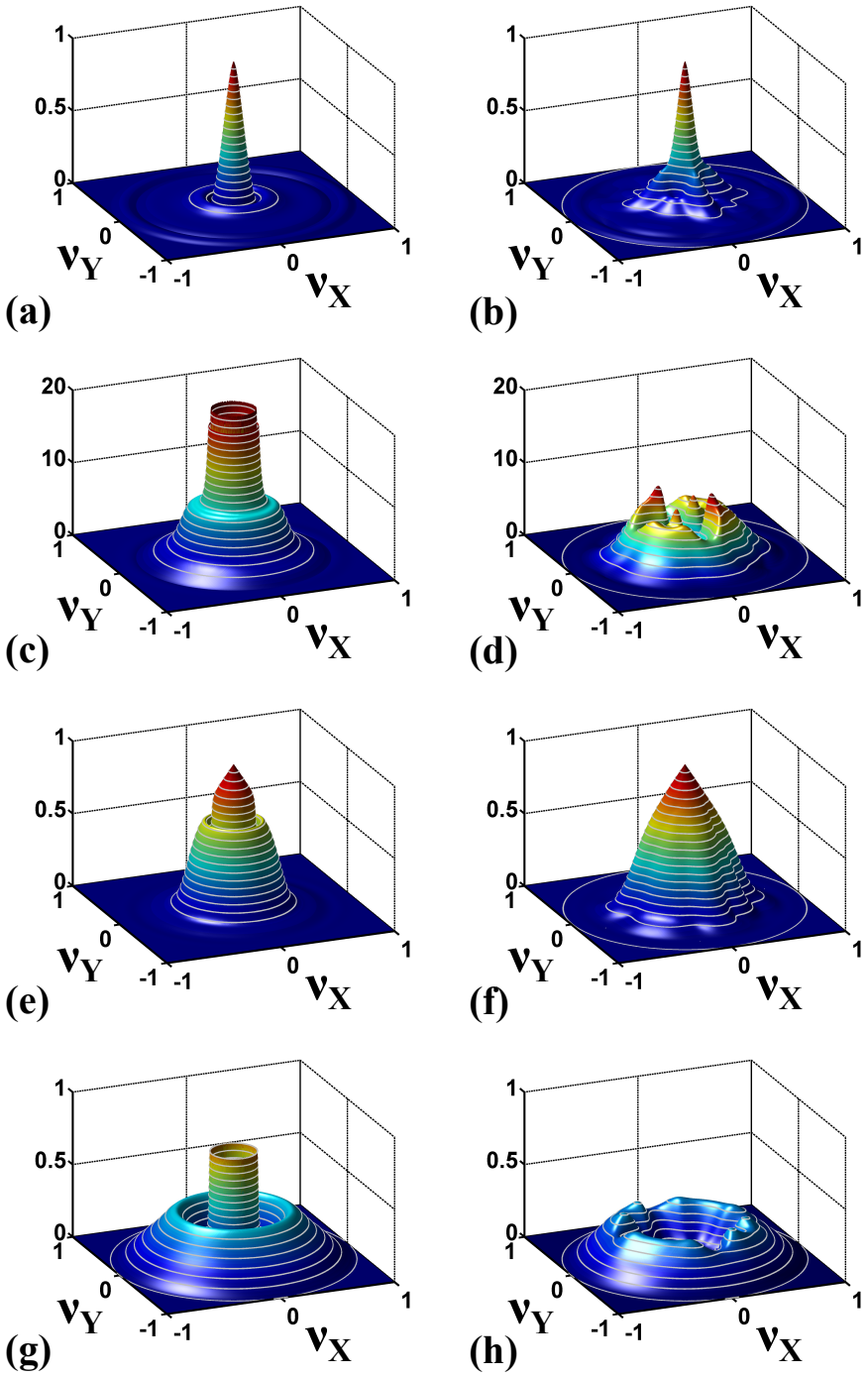


Figure 2.1: *Linear image restoration in the spatial-frequency domain for two optical systems out-of-focus by $W_{20} = 1\lambda$. Left column: a conventional system, right column: a system incorporating a cubic phase-modulation with magnitude $\alpha = \lambda/2$ (defined formally in equation 2.14 on page 30). From top to bottom the plots show the MTF (a-b), the magnitude of the wiener filter (c-d), the restored MTF (e-f), and the magnitude of the difference between the restored OTF and the diffraction-limited OTF (g-h). Note that all plots have the same z-scale and contour lines at $^1/_{20}{}^{th}$ of a unit, except for plots (c) and (d) which have a larger z-scale and contour lines at the unit.*

can be modelled, which after incorporation in equation (2.9) enables the quantification of the expected imaging error of any shift-invariant hybrid imaging system that uses the mean-square-error minimising Wiener filter. The applicability to systems with a spatially varying PSF using more advanced image restoration algorithms will be discussed in section 2.5.2.

2.3 Modelling noise

As for any measuring device, a digital imaging system has a finite precision which introduces a stochastic error commonly referred to as noise. The influence of noise on image quality is even more pertinent for hybrid imaging systems where digital post-processing selectively suppresses or amplifies the noise while correcting for the MTF suppression. In this section an overview of the nature and origin of noise in digital imaging systems is given, as well as a discussion of how it can be modelled practically for the performance prediction of hybrid imaging systems. Detector technology is a very active area of investigation. Instead of discussing the different technologies that are being developed, the following paragraphs give a general description of the different sources and types of noise that are typically found in a digital sensor array.

2.3.1 Analog to Digital Conversion

To allow digital processing, the signal representing the pixel-irradiance must be converted to a digital form, typically having between 8 and 16 bits precision. In the spatial domain it is straightforward to simulate the digitisation process by calculating the fractional number of grey levels that would be filled, and consequently rounding these down to the nearest grey level. However, it is often more practical to use a Fourier representation of the noise.

The Fourier transform of the noise is a linear combination of the digitisation errors of all pixels. Most detectors have a linear response, hence the irradiance ranges are equal in size and equally spaced. In this case it is possible to model the digitisation error as a uniformly distributed random variable with mean zero, and a maximum value of half a grey level. According to the central limit theorem, if the quantisation errors at the pixels are uniform and independent, for a large number of pixels the digitisation error in the spatial-frequency domain is approximated by a normal distribution with an equivalent standard deviation of $1/\sqrt{12}^{\text{th}}$ of a grey level.

Non-uniformity can only be introduced by non-linearity of the detector because the Fourier transform gives equal weight to all the pixel values. Under the assumption that even under non-linearity the irradiance range of each grey level has a finite minimum and maximum length, it can be seen that the Lindeberg condition for convergence to the

normal distribution holds[1]. Independence of the pixel values cannot always be guaranteed, e.g. often some pixels are saturated, causing a bias towards negative digitisation errors. However, for typical detector illumination, simulations show that any deviation from the normal distribution with zero mean is insignificant.

2.3.2 Detection errors

Often a significant error is introduced in the analog part of the detector array. Most detectors, even cooled detectors, have some level of dark signal. If for example the information about the received photons is stored as an electrical charge such as in the case of a CCD array, thermally excited electrons can also get trapped in the potential well before the electronics convert the analog signal to its digital equivalent. It is straightforward to subtract the dark signal from the measured signal to get an on-average correct measurement. However, the uncertainty in the dark signal will increase the measurement error. If the dark current can be considered a Poisson process, then the error it causes, the *dark noise*, will have a variance equal to the mean dark signal in electrons. The dark signal is proportional to the integration time, hence detector specifications usually give the dark current, or the mean dark signal per second.

Also materials used for building the optics and sensor emit thermal radiation. Not surprisingly, thermal infrared detectors will receive many unwanted photons caused by the temperature of objects other than those imaged. Practically this form of noise can also be considered dark noise. Bolometric detection will to some extent also be influenced by thermal conduction. In general, the detector noise of a thermal sensor array is given as the noise equivalent temperature difference (NETD), indicating the standard deviation of the combination of all noise sources in the detection process. In this case it is most practical to model dark noise as a Gaussian distributed random variable.

All detectors can have another source of noise, *read-out noise*, of particular importance in cooled detectors. This error is introduced during the transfer of the detected signal to the analog-to-digital convertor. E.g. in CCD detectors this is caused by electrons that are not removed from the potential-well and are therefore added to one of the neighbouring pixels. In contrast to dark noise, it is independent of the integration time. Read-out noise is typically quoted as a standard deviation, it is therefore most practical to model also this type of noise as Gaussian distributed.

Other errors such as imperfect manufacturing, cosmic rays, or saturated pixels are not specifically considered in this work. Invalid pixels could be isolated and replaced by

[1]The Lindeberg condition can be seen to hold for N irradiance measurements with uniformly distributed analog-to-digital conversion errors if $\forall \xi > 0 \left| \lim_{N\to\infty} \sum_{k=1}^{N} \int_{\min(\xi S_N, r_k/2)}^{r_k/2} x^2 \mathrm{d}x = 0$, where $S_N = \sqrt{\sum_{k=1}^{N} r_k^2/12}$, and r_k is the size of the irradiance range represented by measurement k. If the size of this range is bound as $0 < r_{\min} \leq r_k \leq r_{\max}$, then, for arbitrarily small ξ, a sufficiently large number of measurements, N, will always exist so that the value $S_N \geq \sqrt{N r_{\min}^2/12}$ is larger than $r_{\max}/2\xi$, and the condition holds.

an interpolation of neighbouring pixel values, however to some extent there will always remain an error with a spatially-variant standard deviation.

2.3.3 Shot noise or photon noise

Photon noise, sometimes referred to as shot noise or photon shot noise, is typically dominated by noise from the previously discussed sources. In cooled detectors these types of noise can be reduced to insignificant levels; however, this is not possible for photon shot noise. The quantised nature of electromagnetic radiation can give rise to a variation in the number of detected photons during the integration time. If the detection time of photons can be considered to be independent, this number will be Poisson-distributed. The probability of n photons arriving in a time-interval Δt can then be calculated as $\mathrm{P}(n,\Delta t) = (r\Delta t)^n \exp(-r\Delta t)/n!$, where r is the mean number arriving per time unit. Correlations between photon arrival times can always be found for non-monochromatic light, so in general the number of detected photons is not necessarily Poisson-distributed. However, the coherence time is often sufficiently short so that on average less than one photon is detected during the coherence time. If the coherence time is also significantly smaller than the integration time, then the photon detections can be considered independent events and the detection can be considered a Poisson process.

Images for a photon-noise limited detector can be simulated readily by generating pseudo-random numbers following a Poisson-distribution with a pixel-dependent mean. Note that even though the mean of neighbouring pixels might be correlated, the underlying random processes are independent as discussed in the previous paragraph. Thanks to a property of the Fourier transform, section 2.3.4 shows how this independence permits the accurate approximation of Poisson noise by Gaussian noise in the spatial-frequency domain.

2.3.4 Non-uniform noise in the spatial-frequency domain

As will be discussed in the next section, most scenes have a distinctive non-uniform Fourier spectrum. It would be natural to believe that the Fourier transform of the Poisson noise for such scenes should also be non-uniform, i.e. that it has a spatial-frequency-dependent variance. For a simple sinusoidal scene, the Poisson noise might look as shown by the solid thin red line in figure 2.2a. Contrary to intuition, no extraordinary large values are found in the spectrum of the noise in figure 2.2b. This is fortunate because it simplifies the signal-to-noise model, and more importantly, it agrees with experimental observations. In what follows it will be shown why the noise spectrum is uniform, and furthermore why the noise in the spatial frequency domain can be considered to be Gaussian with zero mean.

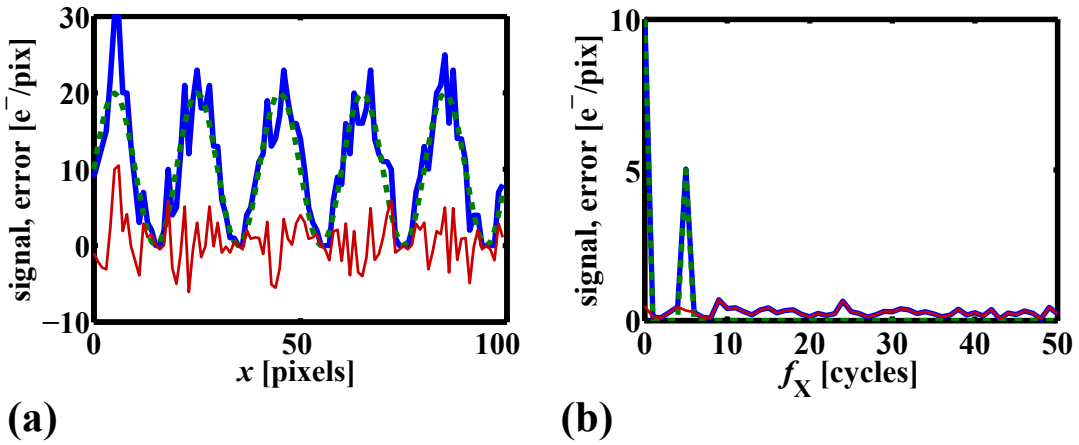


Figure 2.2: *A Poisson-distributed signal (solid thick blue line) with sinusoidal mean (dashed green), and the difference between the real signal and its mean (solid thin red); (a) in the spatial domain plotted as a function of the pixel index x, and (b) in the spatial-frequency domain plotted as a function of the number of cycles over a* 100*-pixel interval. Units are in electrons per pixel, which in (b) is to be interpreted as the spatial mean of the corresponding Fourier component. Note that as the signal is zero for all but two spatial frequencies in (b), the Poisson-distributed signal and the noise coincide for most of the interval.*

The reason behind this lies in the properties of the Fourier transform and in the independence of the noise between pixels. The independence of the noise means that the autocorrelation of the noise must be on average zero for all non-zero shifts, i.e. that it must be proportional to the two-dimensional Dirac delta function[2] when averaged over an infinite number of frames. The power spectrum of the noise is simply the Fourier transform of the autocorrelation, and since the mean of the autocorrelation is proportional to the Dirac delta function, the mean of the power spectrum must be independent of spatial-frequency. In other words, the variance of the noise in the spatial-frequency domain is constant, or the noise is uniform in the frequency domain.

Although the noise between pixels is independent, there must however be a correlation between the noise at different spatial frequencies, otherwise the inverse Fourier transform would result in uniform white noise instead of returning the noisy sine pattern of figure 2.2a. This is less evident, but can be understood knowing that the discrete Fourier transform is a unitary transform[3], i.e. a rotation or mirroring in N-dimensional space, N being the number of pixels. Figure 2.3 demonstrates this for a two-pixel

[2]The Dirac delta function $\delta(x)$ is an asymptotic function, equal to zero for $x \neq 0$ and with indefinite integral equal to 1. The two-dimensional Dirac delta function in the variables x and y is defined as $\delta(x,y) = \delta(x)\delta(y)$.

[3]The continuous Fourier transform is also a unitary transform in the space of complex functions with the L^2 norm. However, the discrete cosine transform is not unitary, independent noise will therefore not necessarily become non-uniform after a discrete cosine transformation.

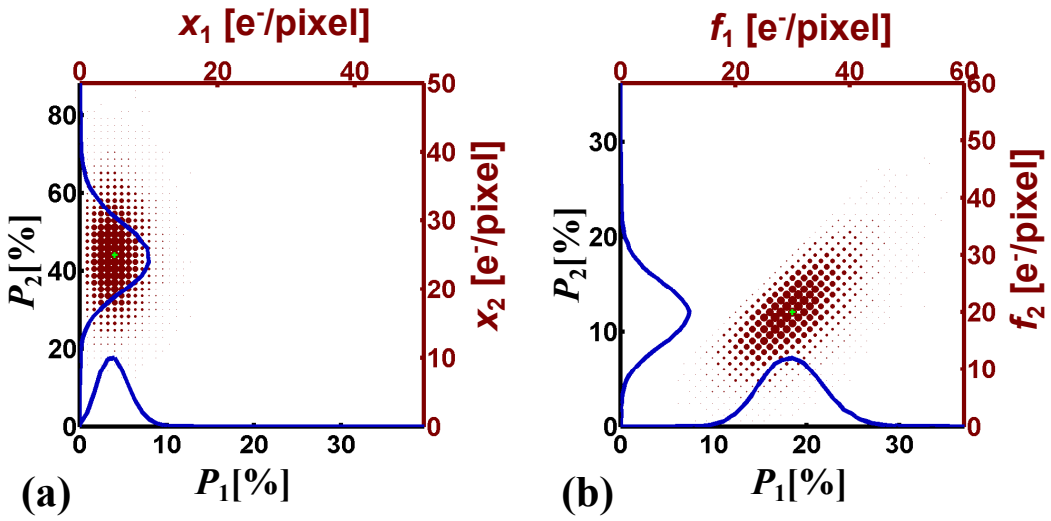


Figure 2.3: *(a) Depiction of Poisson measurement of two variables x_1 and x_2 as would be the case for a two pixel image with mean photo electron count of respectively* 5 *and* 25*. To visualise the distribution,* 100000 *samples are plotted as discs with diameter proportional to the number of occurrences. The mean is marked with a green cross. For each variable separately the probability density functions* $\mathrm{P_x(x_1)}$ *and* $\mathrm{P_y(x_2)}$ *are plotted along their respective axis. (b) The absolute value of the Fourier components of the same samples.*

detector. The photo-electron count for both pixels is Poisson-distributed, however at pixel x_1 with mean 5 $\mathrm{e^-}$, and at pixel x_2 with mean 25 $\mathrm{e^-}$. This leads to the non-rotationally symmetric distribution of measurement values depicted in figure 2.3a, the disc size is proportional to the number of events. The underlying independent probability density functions are plotted along the axes for both pixels. As can be seen in figure 2.3b, The Fourier transform has the property to rotate all axes of the hypercube to a diagonal of the hypercube, leading to the constant variance of the noise after Fourier transformation. The ellipsoidal-like diagonal distribution seen in figure 2.3b is indicative of a correlation between the Fourier components, which can be understood from the non-uniformity of the noise in the spatial domain.

2.3.5 A representative noise model for use in simulations

Following from the arguments presented in the previous subsections, the noise can be modelled accurately in the spatial domain as a combination of Poisson noise and Gaussian noise, or alternatively in the spatial-frequency domain as Gaussian noise. In summary, the recorded pixel value i_{rec} can be simulated as:

$$i_{\mathrm{rec}} = \left\lfloor \frac{\mathrm{Poisson}\left(\iint \phi(x,y) i_{\mathrm{opt}}(x,y)\mathrm{d}x\mathrm{d}y + n_{\mathrm{dark}} T_{\mathrm{int}}\right) + \mathrm{Gaussian}\left(n_{\mathrm{read}}\right)}{N_{\mathrm{e^-,max}}/N_{\mathrm{greylevel}}} \right\rfloor, \quad (2.10)$$

where $i_{\text{opt}}(x,y)$ is the irradiance as a function of the pixel position, $\phi(x,y)$ the detection efficiency defined as zero outside the pixel, and the brackets $\lfloor\cdot\rfloor$ indicate rounding down to the nearest integer. The dark current n_{dark} increases the detected value during the integration time T_{int}, and can be simulated with the same Poisson process if it can be considered independent of the photon noise and both are Poisson-distributed. The read-out noise is often given as a standard deviation n_{read}, and is modelled as a Gaussian distribution in the simulations presented in this chapter. The analog-to-digital conversion is simulated by counting the number of filled grey levels, where each grey level is assumed to represent the same number of detection events: $N_{\text{e}^-,\text{max}}/N_{\text{greylevel}}$. Here $N_{\text{greylevel}}$ is the number of grey levels of the analog-to-digital convertor, and $N_{\text{e}^-,\text{max}}$ is the total number of events required to saturated it. The pixel value has a standard deviation of:

$$\sigma_{\text{n}} = \frac{\sqrt{\iint \phi(x,y) i_{\text{opt}}(x,y)\text{d}x\text{d}y + n_{\text{dark}} T_{\text{int}} + n_{\text{read}}^2}}{N_{\text{e}^-,\text{max}}/N_{\text{greylevel}}} + \frac{1}{\sqrt{12}} \quad [\text{greylevels}], \quad (2.11)$$

or a peak signal-to-noise of:

$$PSNR = N_{\text{greylevels}}/\sigma_{\text{n}} \quad (2.12)$$

$$= \frac{N_{\text{greylevels}}}{\frac{\sqrt{\iint \phi(x,y) i_{\text{opt}}(x,y)\text{d}x\text{d}y + n_{\text{dark}} T_{\text{int}} + n_{\text{read}}^2}}{N_{\text{e}^-,\text{max}}/N_{\text{greylevel}}} + \frac{1}{\sqrt{12}}}. \quad (2.13)$$

Typical values for a cooled CCD detector (Retiga 4000R) are: $n_{\text{dark}} = 1.64\ \text{e}^-/s$, $T_{\text{int}} = 40\ ms$, $n_{\text{read}} = 12\ \text{e}^-$, $N_{\text{greylevels}} = 4096$ (12bits), and $N_{\text{e}^-,\text{max}} = 40\ 000\ \text{e}^-$. Without illumination this amounts to a standard deviation of approximately 1.5 grey levels, which is in agreement with laboratory measurements. Photon noise is dominant for this detector, hence the noise will increase with the illumination level. The noise level will depend on the structure of the scene, however if the detector array is well-illuminated but not saturated, the peak-signal-to-noise level with this detector is experimentally found to be approximately 280 for typical scenes. For this detector, an average illumination of approximately 40% or 16 000 captured photo-electrons would have resulted in a peak-signal-to-noise ratio of 283. Although some variation can be expected in this value, in this chapter a peak-signal-to-noise ratio of 280 is chosen for the calculation of the expected imaging error.

2.4 Modelling the signal

Practical image restoration techniques do not attempt to restore the suppressed contrast to its original level, this would lead to excessive noise amplification. Instead a trade-off between noise reduction and the ideal object restoration is made. This trade-off depends

both on the noise level and on the signal characteristics. As discussed in the previous section, the noise can usually be modelled as Gaussian or Poisson noise, and can be measured or predicted for the majority of systems. To be able to study the performance of image restoration it is also necessary to have some knowledge of the input signal, i.e. the object or scene to be imaged.

The imaged objects will be application dependent, e.g. microscope images might show electronic circuits or biological samples with very different characteristics. Still, most objects of interest have regions of similar intensity and gradients. In this case, points in the same vicinity can be expected to have a significantly correlation in intensity, decreasing with the distance between them. If a universal correlation exists for all natural scenes, it should be the same for magnified parts of the same scene, hence it should be shift- and scale-invariant. If such a correlation exists between the pixels and assuming a maximum entropy probability density, it is possible to deduce that the expected power spectrum[4] of images should decrease as a power of the spatial-frequency [84, 85]. Although these assumptions certainly will not always hold, the proposed model does provide a potential explanation of why numerical analysis of large image sets consistently finds such a power law [86–88], with an exponent κ between -1 and -1.5 found by most empirical studies. Although scale- and shift-invariance should lead to such a power law, it does not provide a value for the exponent.

An intuitive explanation for the experimentally found exponent value could be given by considering that natural scenes consist mainly of areas of similar intensity bordered by edges where the intensity changes abruptly. Without loss of generality it is possible to consider a line profile of the scene. Such a section could be modelled as a linear combination of scaled and translated Heaviside unit step functions $H_{\text{step}}(x)$, $I(x) = \sum_k a_{\text{k}}(2H_{\text{step}}(x - b_{\text{k}}) - 1)$, where the constants a_{k} and b_{k} determine respectively the amplitudes and positions of the intensity changes. The Fourier transform of the terms is $\mathscr{F}\{I(x)\}_f = \sum_k a_{\text{k}} \text{i}\sqrt{2/\pi}\exp(\text{i}b_{\text{k}}f)\, f^{-1}$. Note that the magnitude of this last expression is also a decreasing power function as that predicted by the model proposed in [84] and [85]. If the amplitudes a_{k} and the positions b_{k} are independent and randomly distributed, then the magnitude of the Fourier transform of the scene intensity will decrease with the inverse of the spatial-frequency, in other words the exponent should have the value -1.

However, large sets of natural scene image spectra have been studied and fit the power function $\|\boldsymbol{\nu}\|^{\kappa}$ well with $\kappa < -1$. Billock compared several previous studies and calculated their weighted average as $\kappa = -1.08$ [86], clearly below the -1 deduced in the previous paragraph. This discrepancy could by due to the optical filtering of the images, i.e. approximately a multiplication in the spatial-frequency domain by $1 -$

[4]The power spectrum of a function is defined as the squared absolute value of the Fourier transform of that function.

$\|\boldsymbol{\nu}\|$. In figure 2.4 an example is shown of both the spatial-frequency distribution before imaging and after imaging with a finite circular aperture. A similar reduction towards high spatial frequencies would occur due to the finite size of the sensor pixels, this will be discussed in more detail in section 3.5 of the next chapter. Although the assumptions made in the previous paragraphs only permit speculation, it could explain the somewhat lower κ values found by many studies.

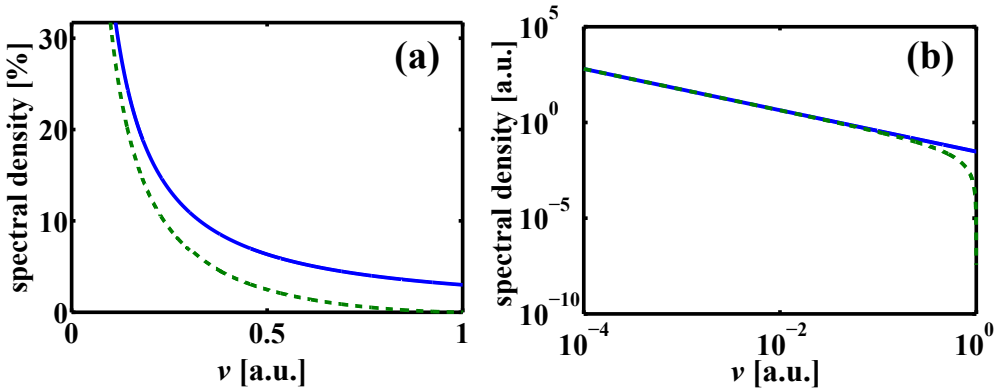


Figure 2.4: *(a) The spatial-frequency distribution of a typical scene follows approximately the solid blue curve:* $0.03|\nu|^{-1.08}$. *Diffraction at a finite circular aperture suppresses the high spatial frequencies as indicated by the dashed green line. (b) The same distribution plotted on logarithmic scale.*

In combination with the noise model proposed in the previous section, an estimate can be made of the imaging error given by equation (2.9). This is illustrated in figure 2.5 for the two out-of-focus imaging systems discussed in section 2.2. The square root of the first term, the regularisation error, is shown as a function of normalised spatial-frequency in the top row. The plots in figure 2.5a and b are in fact the weighted equivalent of those in respectively figure 2.1g and h. The square root of the second term is shown on the plots in the second row, figure 2.5c and d are proportional to those in respectively figure 2.1c and d. A combination of the plots in the first two rows gives the total imaging error ε as shown in the bottom row. The plots in figure 2.5e and f are thus a weighted version of a linear combination of respectively the plots in figure 2.1c and g, and those in figure 2.1d and h. When comparing the plot in the top row with those in the bottom row, it can be noticed that the noise amplification is only partially responsible for total imaging error. The regularisation used in the image restoration depends on the signal-to-noise ratio, and therefore the noise level has also an indirect, but important influence on the final imaging error. In this example, the signal level was chosen so that the signal-to-noise ratio is exactly $^{10}/_{\boldsymbol{\nu}}$, a ratio representative for typical scenes imaged with a well-illuminated photon-limited detector. The signal-to-noise ratio at $\|\boldsymbol{\nu}\| = {}^1/_2$ is thus approximately $10\left(4\pi - 3\sqrt{3}\right)/3\pi \approx 7.82$ for a circular aperture.

It can be noted that the imaging error in figure 2.5e has significantly larger values compared to those seen in figure 2.5f. While the introduction of a cubic phase-

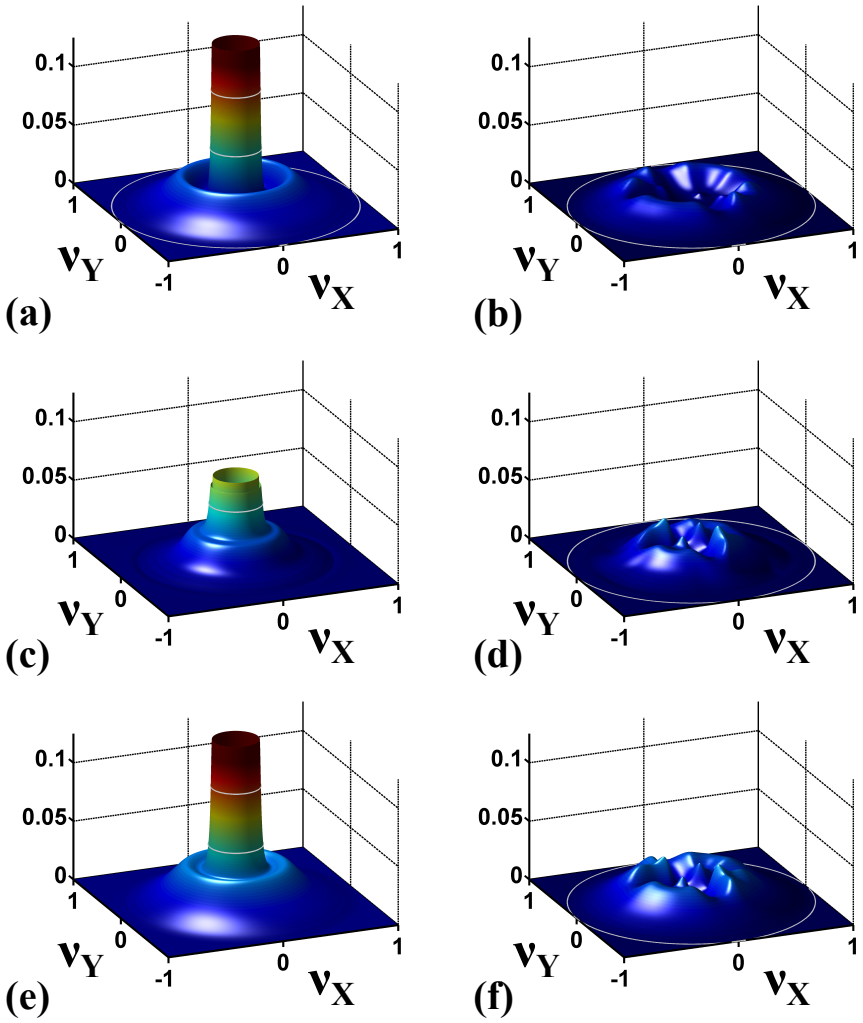


Figure 2.5: *The regularisation error (a-b), the error due to noise amplification (c-d), and the total imaging error ε (e-f), as given by equation (2.9). The plots on the left correspond to a system without pupil modulation out-of-focus by $W_{20} = 1\lambda$, and the plots on the left correspond to the same out-of-focus system after the introduction of a cubic pupil-phase modulation with $\alpha = \lambda/2$.*

modulation reduced the MTF, sufficiently high contrast is maintained so that overly large imaging errors can be avoided. This could already be noticed in figure 2.1, however the relative importance of low spatial frequencies makes the difference between the two designs more obvious.

Although the signal diverges for $\boldsymbol{\nu} \to \mathbf{0}$, this is merely a technical problem because at this spatial-frequency the detected signal represents the total irradiance which will be unaffected by the Wiener filter when the signal-to-noise ratio becomes infinite. The regularisation error represented by the first term in equation (2.9) should thus have no contribution for spatial-frequency zero. Numerical problems can be avoided by setting the regularisation error to zero for this spatial-frequency, or by capping the signal model to the maximum irradiance that can be measured with the detector array.

2.5 Evaluation of the defocus tolerance of common phase-modulations

2.5.1 Introduction

The use of pupil-plane phase-modulation in hybrid optical-digital imaging systems can enable good imaging quality even in the presence of high levels of optical aberration, such as defocus, astigmatism and spherical aberration [18, 89, 60, 52]. The method introduced in the previous sections is now applied to find the phase-profile and modulation depth that maximise the expected image-fidelity for a required aberration tolerance. Crucial for such a hybrid design is a well-considered choice of phase-profile used for optical encoding. We describe here the optimisation of the most important phase-modulations for reduced sensitivity to defocus; namely third-order antisymmetric profiles and radially-symmetric quartic phase-modulation. It should be noted however that the technique is more generally applicable to higher order aberrations, as well as to other phase- or amplitude-modulations.

Various types of phase-modulations have been shown to increase tolerance to defocus; one of the most common is the cubic phase-modulation and its generalisations of the form:

$$\theta(x,y) = \alpha\left(x^3+y^3\right)+\beta\left(x^2y+xy^2\right), \tag{2.14}$$

where x and y are normalised pupil coordinates, and the parameters α and β characterise the phase-modulation. Phase-masks such as the separable logarithmic [63], the exponential [61] and the generic polynomial phase-mask [62] have profiles similar to the pure-cubic phase-profile; that is, with $\beta = 0$. Also non-separable profiles, such as those described by equation (2.14) with $\beta \approx -3\alpha$ have been found to yield systems with good defocus tolerance [64, 19].

Also of interest are radially-symmetric phase profiles such as the logarithmic asphere [50], the quartic phase-mask [56, 52, 60] and spherical coding [90, 58, 59]. The latter two can be described as:

$$\theta(\rho) = \gamma\left(\rho^2+\delta\right)^2, \tag{2.15}$$

where ρ is the normalised radius, and the parameters γ and δ characterise the phase-modulation.

In section 2.5.3, hybrid optical-digital imaging systems combining different types of pupil-phase modulation with optimal image restoration are evaluated. The applicability of the proposed image-fidelity metric to more advanced image restoration techniques that enable near-optimal image restoration is discussed in the following section.

2.5.2 Optimal image restoration

Various assumption have been made in the discussion of section 2.2 to derive the optimal image restoration. Although the widespread use of the Wiener filter shows that these assumptions are often practical, more advanced image restoration algorithms exist that do not require spatial invariance or exact knowledge of the PSF and signal-to-noise ratio. Such algorithms cannot be represented in the concise Fourier-space notation as used until now, and often the restored image can only be approximated using an iterative procedure. The typically longer calculation time is however justified by the more accurate image restorations that can be obtained by making stronger assumptions such as piecewise smoothness of the scene, or non-negativity[5] of the scene irradiance, and PSF values. Although disregarded by the Wiener filter, the absolute certainty of the non-negativity constraint provides extra information that can be used to calculate a restored image that is a closer approximation of the undistorted noiseless image. This is exploited by the *Iterative Constraint Tikhonov Method* or ICTM [91], and indirectly also by the Richardson-Lucy algorithm [92, 93] which assumes that the measured value is a Poisson-distributed variable with mean proportional to the irradiance.

Optics designed for an aberration-tolerant hybrid imaging system have, in principle, an aberration-invariant PSF. It could thus be argued that there is no significant benefit in using these more computationally intensive algorithms. However, small changes in the PSF are unavoidable in practice, and can cause artifacts in the restored image. It could be for example that there is a phase-mismatch between the restoration filter and the actual OTF. This will lead to what appears to be edge replications in the image [47], for an image of an optical system with cubic phase-modulation that has been restored with the kernel corresponding to an incorrect distance. Although the PSF of the optical system could be made arbitrarily defocus-invariant, this would lead to excessive noise amplification as will be shown in the next subsection. A trade-off will therefore be required to obtain the optimal hybrid imaging quality.

New image restoration algorithms are being published regularly, in particular for blind image restoration that deconvolves with an unknown, or partially unknown PSF. The aforementioned constraints permit blind deconvolution algorithms to obtain an accurate estimate of the PSF and the restoration. Although the tighter constraints can counteract the lack of information, noise still imposes a fundamental limit on the achievable image quality. To be able to estimate the performance of a hybrid imaging system relying on such image restoration techniques, the discussion in this chapter assumes that the optimal performance is achievable by use of the mean-square-error minimising

[5] Non-negativity of the restored image cannot always be assumed when the pixel values are not directly proportional to irradiance, e.g. for a thermal imaging system where relative temperature is measured. In this case the constraint that the restored values should not be below 0 K will probably provide little additional information.

restoration in which the aberration is known *a priori*. That is; the kernel used in image restoration is optimally matched to the actual PSF for each level of aberration. This is appropriate when the aberration varies in a known way, e.g. with field position [94], or with zoom [95]. Even without *a priori* knowledge of the PSF, near-optimal image recovery is possible using blind-deconvolution techniques [96, 97]. For third-order phase modulations it has been shown that depth can be estimated accurately from defocus [98], thereby potentially increasing the convergence speed of blind-deconvolution algorithms. In general it can thus be expected that the optimal image restoration can be approached using advanced image restoration techniques.

2.5.3 Defocus tolerance with third-order phase-modulation

The importance of the influence of the phase-modulation amplitude on attainable image quality is illustrated by the simulations shown in figures 2.6 and 2.7, which depict the variation of image error with defocus and amplitude of phase-modulation, α, for a pure-cubic phase function. The top row shows the error magnitudes of simulated images obtained with a hybrid imaging system incorporating a sub-optimal phase-modulation, $\alpha = {}^1\!/_2\alpha_{\text{opt}}$, for defocuses of $W_{20} = 0\lambda$, $W_{20} = 3\lambda$, and $W_{20} = 5\lambda$ subsequent to restoration with a Wiener filter. The optimal phase-modulation, $\alpha_{\text{opt}} = 2.87\lambda$ is calculated using equation (2.2) for a defocus tolerance, $|W_{20}| \leq 5\lambda$. The error magnitudes are calculated with respect to the ideal, noiseless, diffraction-limited images. Error magnitudes for systems incorporating the optimal, and twice the optimal modulation depth are shown respectively on the following two rows.

It can be appreciated that for small values of α the suppression of the modulation-transfer function for larger values of defocus introduces increasing error magnitudes in the recovered images (see for example the image for $\alpha = {}^1\!/_2\alpha_{\text{opt}}$, $W_{20} = 5\lambda$). For increasing values of α, the errors become less sensitive to variations in W_{20}; however for $\alpha > \alpha_{\text{opt}}$, the error averaged across the range of defocus reaches a minimum and starts to increase due to the increased suppression of the MTF by the pupil-plane phase-modulation. A large number of such simulations for representative scenes enables identification of the optimal α for these particular scenes, however, it is possible to obtain a more general result by assuming a model for the scene spatial spectrum, as suggested in section 2.4.

By way of example, equation (2.9) is used to estimate the expected imaging-error, $\bar{\varepsilon}$ with a signal-to-noise model: ${}^{P_S(\|\boldsymbol{\nu}\|)}\!/_{P_N} = \left|{}^{10}\!/_{\|\boldsymbol{\nu}\|}\right|^2$, that is $\kappa = -1$, and with the standard deviation of the noise equal to ${}^1\!/_{280}{}^{\text{th}}$ of the dynamic range. These are typical values for a well-illuminated, high dynamic range detector operating in detector-noise-limited conditions. In figure 2.8a, ε is plotted as a function of W_{20} for several values of α. It can be appreciated from this figure that larger values of α yield larger values of ε for

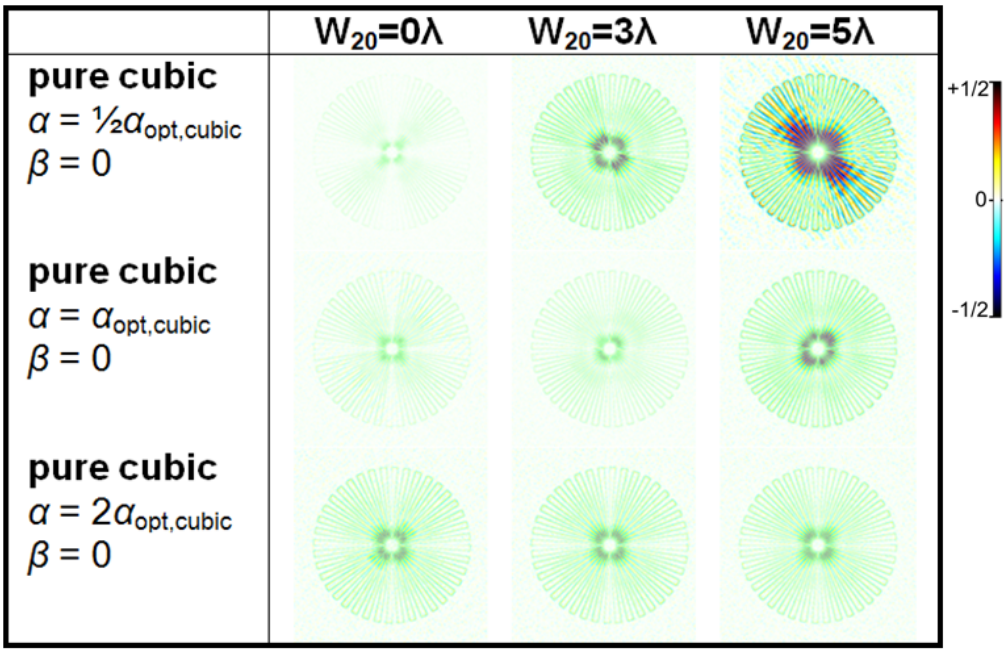


Figure 2.6: *False-colour coding of error magnitudes of simulated spoke targets recorded with pure-cubic pupil-phase modulation ($\beta \equiv 0$) for various values of α and W_{20}. The colours indicate pixel value differences between the restored image and the ideal, noise-less diffraction-limited image, normalised to the dynamic range. The optimal modulation depth, $\alpha_{opt,cubic}$ of the pure-cubic for a defocus tolerance $|W_{20}| \leq 5\lambda$ is 2.87λ. Notice that the green hues are indicative of the colour mixing due to the proximity of positive (yellow) and negative (blue) errors.*

small W_{20} (due to increased suppression of the MTF). However, ε is then less sensitive to defocus; hence for larger W_{20}, a larger α yields a smaller ε (by avoiding the zeros in the out-of-focus MTF of a traditional optical system). In applications for which a finite range of defocus up to a certain maximum value $W_{20,\max}$ is encountered, then $\bar{\varepsilon}$, the quadratic mean of ε over the range of W_{20} is a useful figure of merit. The solid-blue trace in figure 2.8b shows the variation of $\bar{\varepsilon}$, averaged over the range $-5\lambda \leq W_{20} \leq 5\lambda$ as a function of α. For this range of W_{20}, the optimal value of α is found to be 2.87λ.

It is interesting to compare the optimal values of $\bar{\varepsilon}$ obtained with the pure-cubic ($\beta \equiv 0$ in equation (2.14), the generalised-cubic masks ($\beta \equiv -3\alpha$), and the radial quartic profile given by equation (2.15). In figure 2.8b, $\bar{\varepsilon}$ is plotted as a function of α for the former two, and in terms of γ for the latter. It can be seen that for these specific parameters minimum values of $\bar{\varepsilon}$ are lowest for the cubic phase profile and are 12% higher for the generalised-cubic and 30% higher for the radially-symmetric mask. Furthermore, simulations have shown that these differences are relatively insensitive to defocus range.

Using the general model, it is straightforward to predict the imaging fidelity of a hybrid imaging system with arbitrary phase- or amplitude-modulation. Here the same analysis is applied to the generalised-cubic phase-profiles described by equation (2.14) to

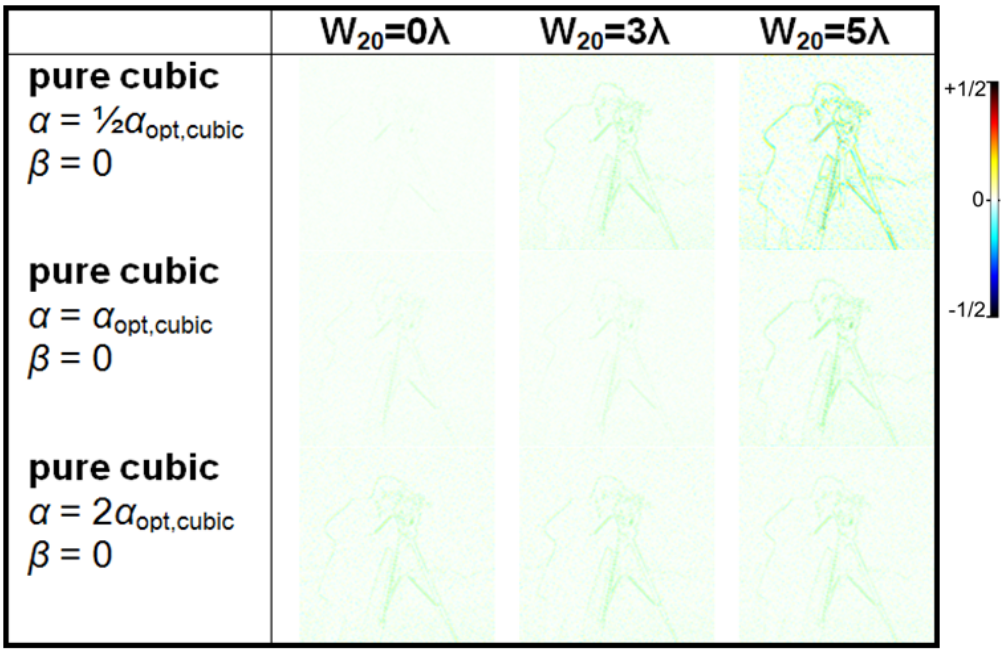


Figure 2.7: *False-colour coding of error magnitudes of simulated images recorded with pure-cubic pupil-phase modulation ($\beta \equiv 0$) for various values of α and W_{20}. The colours indicate pixel value differences between the restored image and the ideal, noise-less diffraction-limited image, normalised to the dynamic range. The optimal modulation depth, $\alpha_{opt,cubic}$ of the pure-cubic for a defocus tolerance $|W_{20}| \leq 5\lambda$ is 2.87λ. Notice that the green hues are indicative of the colour mixing due to the proximity of positive (yellow) and negative (blue) errors.*

determine the optimum values of α and β independently. The contour plot in figure 2.9 shows the variation in imaging error, $\bar{\varepsilon}$, as a function of α and β for a system with a defocus tolerance $|W_{20}| \leq 5\lambda$. A global optimum imaging fidelity is obtained near the axis $\beta \equiv 0$, and a secondary maximum can be noted on the axis $\beta \equiv -3\alpha$ in figure 2.9.

Both optimal sets of (α,β) are plotted in figure 2.10a as a function of the maximum value of defocus to be mitigated, $W_{20,\max}$. A quasi-linear increase of the global optimum of $\alpha \approx {}^1\!/_2 W_{20,\max}$ can be observed. This is in agreement with the analytically derived expression for the optimum value of $W_{20,\max} = (1 - \|\boldsymbol{\nu}\|)\, 3\alpha$ given in [99] when employing an effective spatial-frequency $\|\boldsymbol{\nu}\| = {}^1\!/_3$. A small deviation of 0.3λ towards larger α-values can be noticed, and the linear dependence breaks down for small defocus tolerances of $W_{20,\max} \approx {}^\lambda\!/_2$. For defocus smaller than this limit, a traditional system with digital image restoration outperforms a system with cubic pupil-phase modulation. The calculations also show that β is optimal at approximately 0.3λ, largely independent of $W_{20,\max}$, although its effect on the performance is essential negligible.

Optimal sets of α and β for the secondary optimum are shown in figure 2.10b and have an approximately linear dependency with $W_{20,\max}$. Notice that the dependency between the two parameters closely follows $\beta \equiv -3\alpha$, corresponding to a trefoil wave-

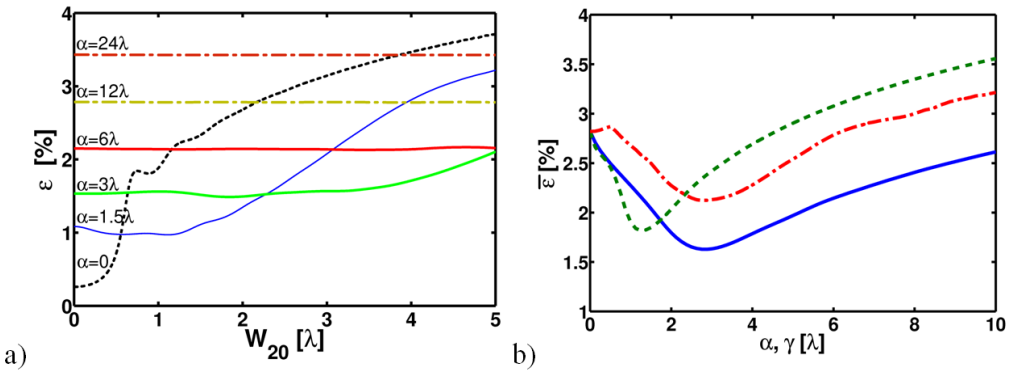


Figure 2.8: *a) The square root of the expected mean-square imaging-error normalised to the dynamic range, ε, is shown as a function of defocus for various cubic phase-modulation depths α; b) shows in solid-blue line the average imaging-error as a function of α for $|W_{20}| \leq 5\lambda$. For comparison, the dashed-green line and dot-dashed-red line show the average imaging error of respectively the generalised-cubic mask with $\beta \equiv -3\alpha$, and the quartic phase-modulation with optimal $\delta = -0.7$, as a function of their respective parameters α and γ.*

front modulation. A similar phase-modulation has been derived by minimisation of the Fisher information in the PSF, while constraining reduction in the Strehl ratio [64, 19].

In figure 2.10c the variation of α is shown with $W_{20,\max}$ for both, approximately pure-cubic ($\beta \approx 0.3\lambda$) and generalised-cubic ($\beta \approx -3\alpha$) phase functions. It can be seen that $\bar{\varepsilon}$ increases monotonically with $W_{20,\max}$ in both cases and that the approximately cubic function provides a slightly lower $\bar{\varepsilon}$. The improvement offered by the approximately cubic mask is modest however and additional criteria may have a bearing on the preferred modulation function. For example, it has been found that high levels of astigmatism in a singlet lens were better mitigated using the trefoil form of the generalised-cubic phase profile [94].

2.6 Perception of hybrid imaging quality

To illustrate the impact of the variation of image error with W_{20}, figures 2.6, 2.7, 2.11, and 2.12 show false-colouring of the image errors for three defocus positions and two scenes: a spoke and a representative scene. Whereas figures 2.6 and 2.7 depict image errors for the pure-cubic phase-modulation ($\beta \equiv 0$), figures 2.11 and 2.12 depict image errors for the generalised-cubic with $\beta \equiv -3\alpha$, and quartic phase functions. Image errors are depicted for $W_{20} = 0$, 3λ and 5λ and for the amplitudes of phase-modulation equal to half the optimal values, the exact optimal values and twice the optimal values, where the optimal amplitudes are $\alpha_{opt,gen} = 1.29\lambda$ for the generalised-cubic and $\gamma_{\text{opt}} = 2.8\lambda$, $\delta_{\text{opt}} = -0.7$ for the quartic mask.

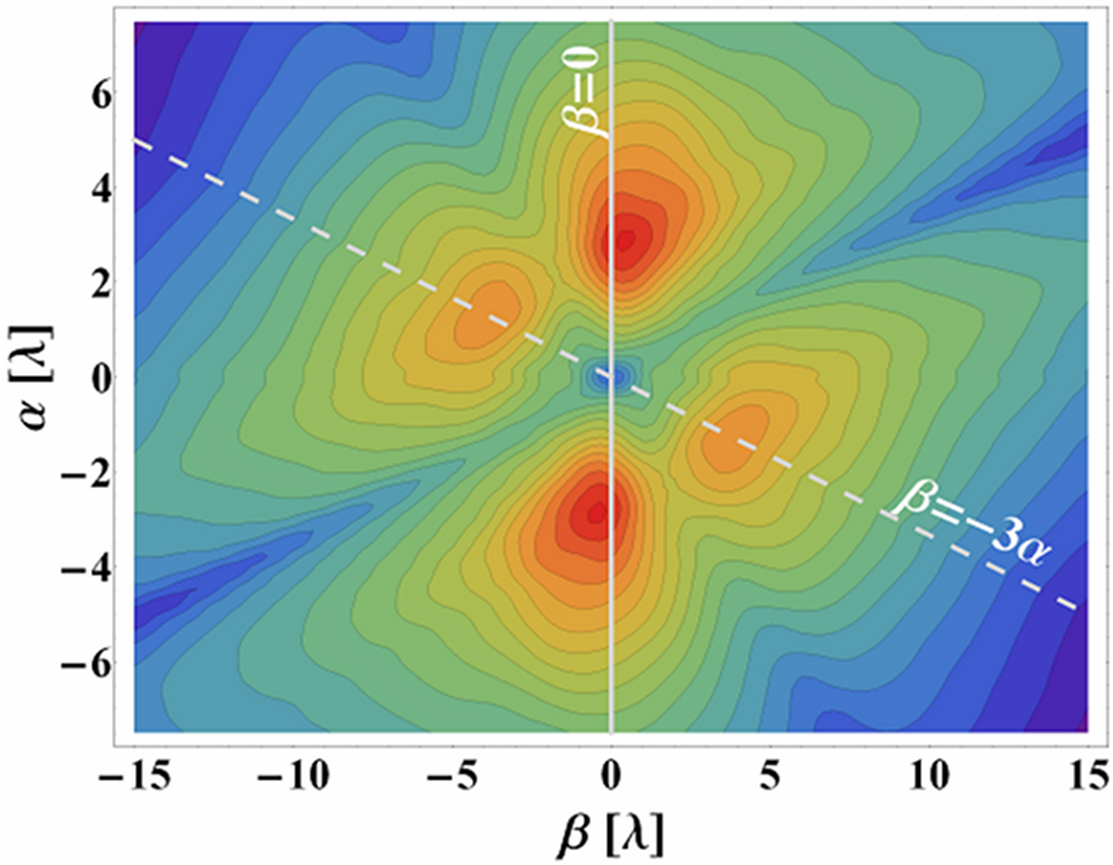


Figure 2.9: *The expected imaging error $\bar{\varepsilon}$ [dB] calculated using a statistical model of the scene, as a function of the parameters α and β for the generalised-cubic phase-modulation defined by equation (2.14).*

As expected and as can be appreciated from these images, the lowest imaging error occurs for zero defocus and lower amplitudes of phase-modulation (images in first column and first and third rows), however for larger defocus (third column) the errors are larger. Similarly larger amplitudes of phase-modulation (third and sixth rows) yields less variation with W_{20}, nevertheless the average values of the errors are larger. Note that only images for positive defocus are shown. The PSF of antisymmetric phase-modulation varies as an even function of defocus; images for negative defocus are therefore identical to images for positive defocus. The same is not true for the quartic phase-modulation for which simulations were conducted for both positive and negative defocus. In this case however, the images shown are representative because the optimised δ-parameter in equation (2.15) shifts the region of low imaging error to coincide optimally with the defocus tolerance range $-5\lambda \leq W_{20} \leq 5\lambda$.

In the images on the second row of figure 2.6 it can be seen that the pure-cubic imaging error is relatively low for the horizontal and vertical spatial frequencies, this is directly related to the high MTF values of the pure-cubic along these axes. The same angular variation is not visible for the images of the generalised-cubic phase function

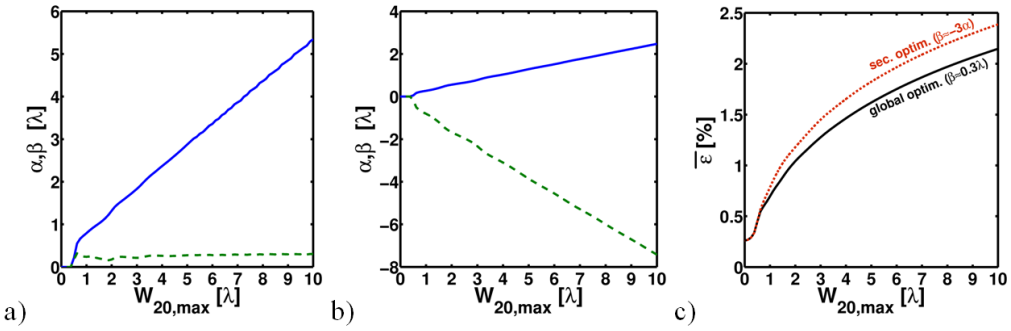


Figure 2.10: *(a) The global optimal phase-modulation parameters α and β as a function of the required defocus tolerance. (b) The secondary optimal phase-modulation parameters as a function of the required defocus tolerance. (c) The root of the expected mean-square imaging-error for both profile types as a function of the maximum defocus $W_{20,\mathrm{max}}$.*

shown on the second row of figure 2.11, however a modest hexagonal variation is still present. The quartic modulation, due to its radial symmetry, has an angle-independent imaging error.

To assess the relevance of the conclusions of section 2.5.3 to perceived image quality, the mean structural similarity metric (MSSIM) has been used to assess image quality. This metric has been shown to correlate well with human perception of image similarity [75]. Instead of using a SNR-model, the metric is applied to a set of commonly used natural and synthetic test-images (figure 2.13b) modified by the corresponding OTFs for 21 defocus positions $-5\lambda \leq W_{20} \leq 5\lambda$. Photon-noise, Gaussian detector-noise, and analog-to-digital conversion at the detector are included in the simulation. The noisy images are then restored using the corresponding Wiener filter, compared to the theoretical diffraction-limited image using the MSSIM metric, and plotted in figure 2.13a. The MSSIM contour plot and the plot in figure 2.9, obtained using the scene-model are very similar with almost identical optimal values of α and β, indicating that the use of equation (2.2) as a metric of image quality yields a valid optimisation of perceived image quality.

2.6.1 Fine-tuning of the metric for the human visual system

As any aperture, the pupil of the human eye works as an optical filter only permitting spatial frequencies below a certain cut-off spatial-frequency f_{cutoff}. The modulation transfer function of such an optical system with circular aperture can be described by the following equation [45]:

$$MTF(\boldsymbol{\nu}) = \frac{2}{\pi}\left(\arccos\|\boldsymbol{\nu}\| - \|\boldsymbol{\nu}\|\sqrt{1-\|\boldsymbol{\nu}\|^2}\right), \|\boldsymbol{\nu}\| \leq 1, \tag{2.16}$$

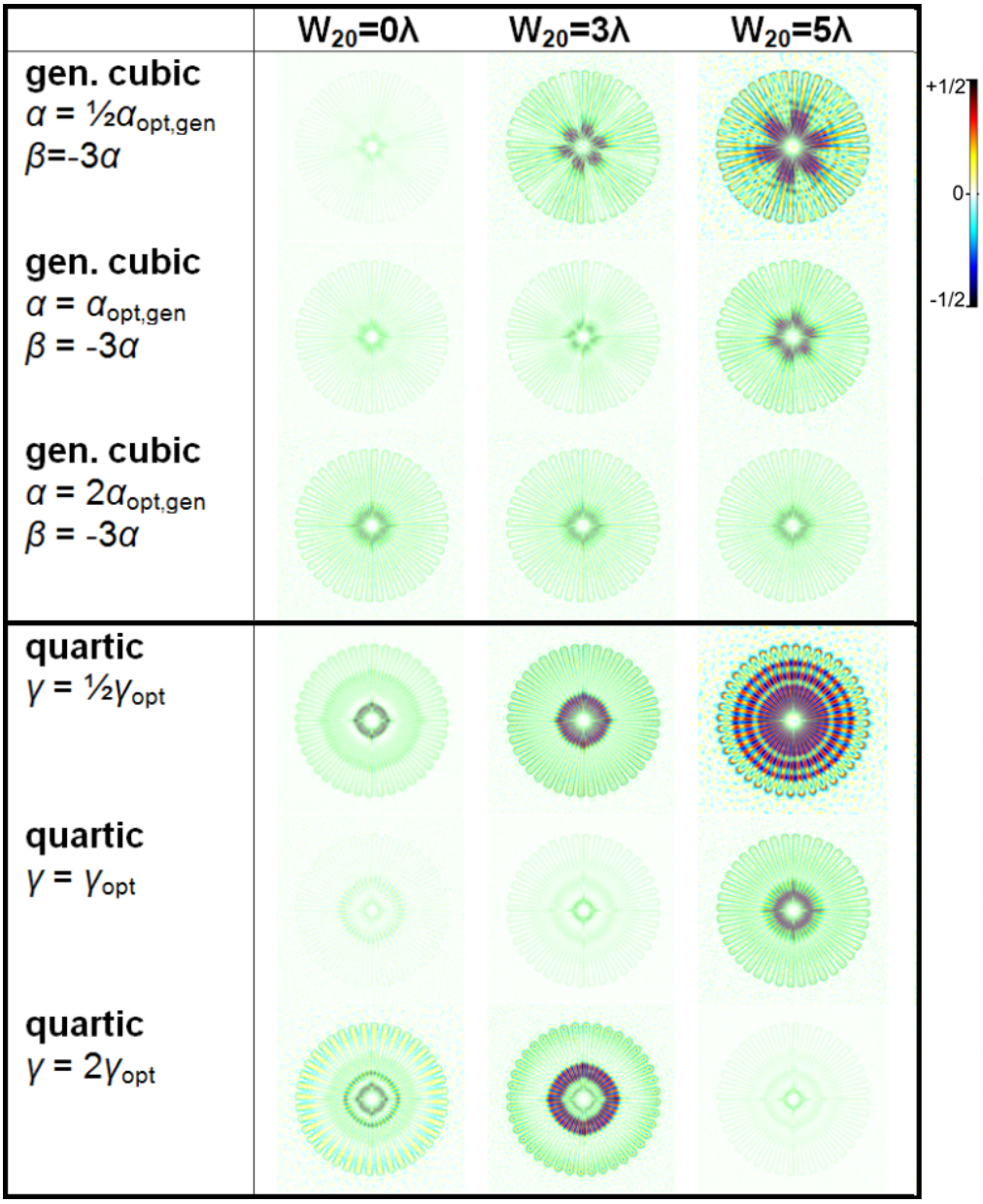


Figure 2.11: *False-colour hybrid imaging error for various phase-modulations and defocus. The secondary optimum and multiples of the generalised-cubic phase-modulation ($\alpha_{opt,gen} = 1.29\lambda$, $\beta = -3\alpha$), are used for the error magnitudes shown in the top three rows. The quartic with optimal parameters $\gamma_{\text{opt}} = 2.8\lambda$ and $\delta_{\text{opt}} = -0.7$ and multiples are used for the error magnitudes in the bottom three rows. The spoke target image at the left in figure 2.13b is used as test-scene. Notice that the green hues are indicative of the colour mixing due to the proximity of positive (yellow) and negative (blue) errors.*

where $\boldsymbol{\nu}$ is the ratio of the spatial-frequency f and the cut-off spatial-frequency, determined by the wavelength, the pupil-diameter, and the focal length of the eye. For most purposes this is an accurate model of the image formation on the retina, however psychophysical tests show that the human ability to distinguish periodic patterns decreases towards the lowest spatial frequencies. This reduced sensitivity is related to the neural

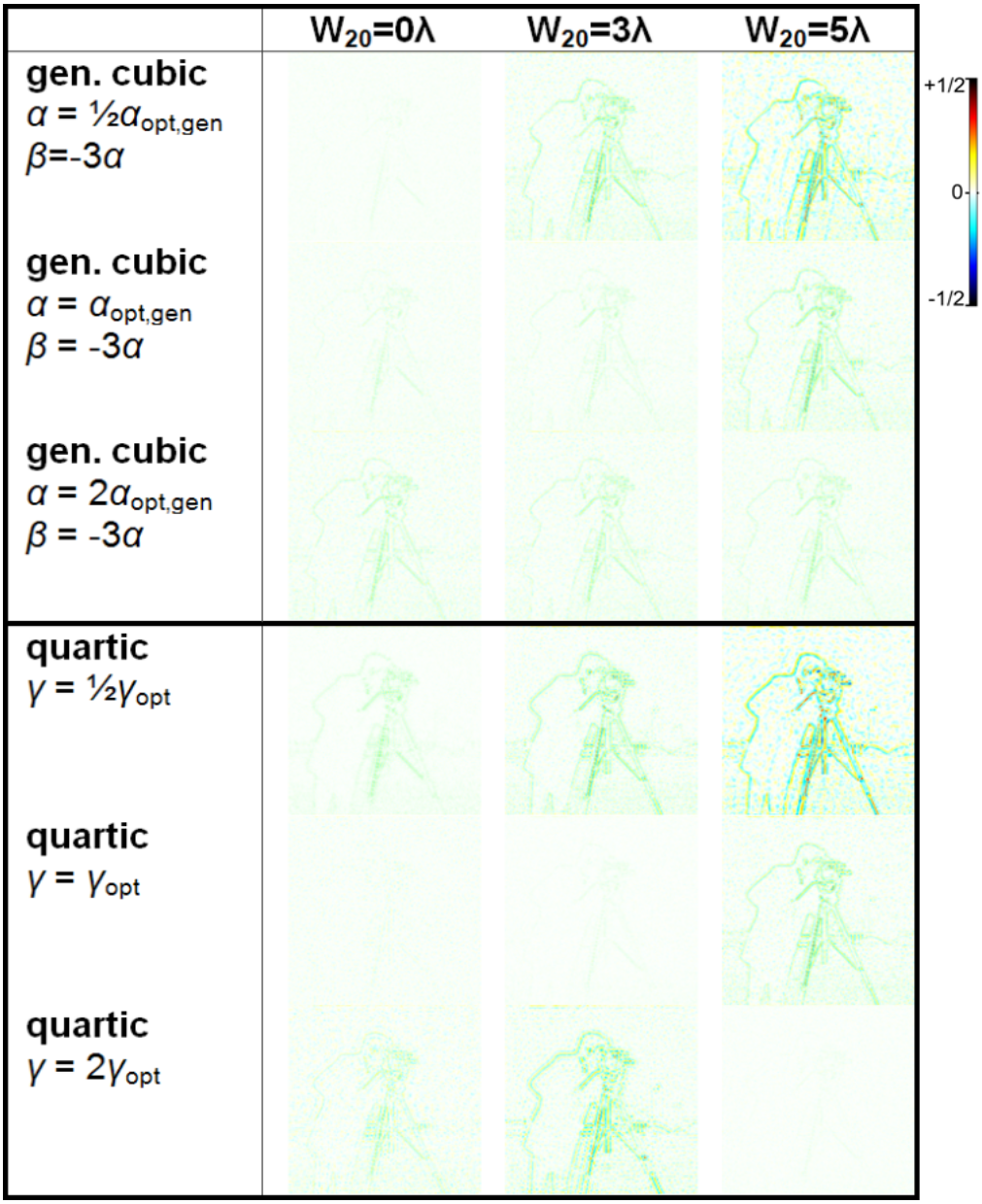


Figure 2.12: *False-colour hybrid imaging error for various phase-modulations and defocus. The secondary optimum and multiples of the generalised-cubic phase-modulation ($\alpha_{opt,gen} = 1.29\lambda$, $\beta = -3\alpha$), are used for the error magnitudes shown in the top three rows. The quartic with optimal parameters $\gamma_{\text{opt}} = 2.8\lambda$ and $\delta_{\text{opt}} = -0.7$ and multiples are used for the error magnitudes in the bottom three rows. The cameraman image, second from the left in figure 2.13b, is used as a test-scene. Notice that the green hues are indicative of the colour mixing due to the proximity of positive (yellow) and negative (blue) errors.*

processing after the image formation. Neurons in the retina can be activated by photon absorption after which the signal is relayed to the visual cortex. At the same time, activated neurons inhibit activation of neighbouring neurons. The combined effect of the crosstalk is akin to an edge detection filter, effectively suppressing the lowest spatial

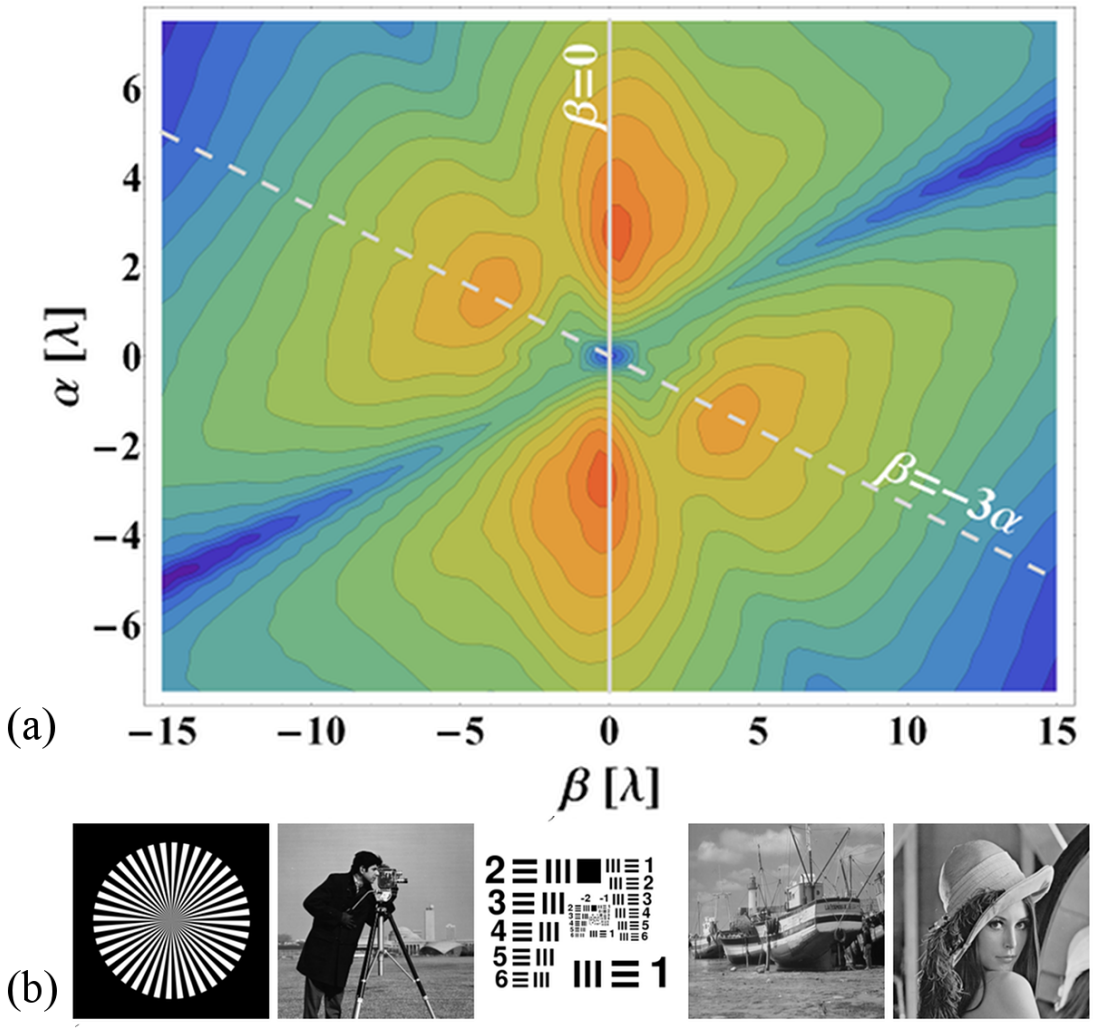


Figure 2.13: *(a) The structural similarity of the same hybrid imaging systems shown in figure 2.9, averaged for the test-scenes shown in (b), retrieved from the SIPI image database (University of Southern California).*

frequencies. Considering that low spatial frequencies tend to dominate in natural scenes (cf. section 2.4), the information-loss is however modest, and probably would have been far greater without the crosstalk due to saturation of the neural signal in bright areas.

Based on psychophysical experimental data Mannos and Sakrison [100] proposed to model the normalised contrast sensitivity function as:

$$CSF(f) \approx 2.6(0.0192+0.114f)\exp\left(-(0.114f)^{1.1}\right), \qquad (2.17)$$

which has a peak around the spatial-frequency $f \approx 8$ cycles/degree as shown in figure 2.14. Contrast sensitivity is defined as the inverse of the minimum perceptible contrast. A more general model accounting for both the spatial and temporal variation of the contrast sensitivity is provided by van den Branden Lambrecht [101]. Although the response of the human visual system also depends on other factors such as wavelength,

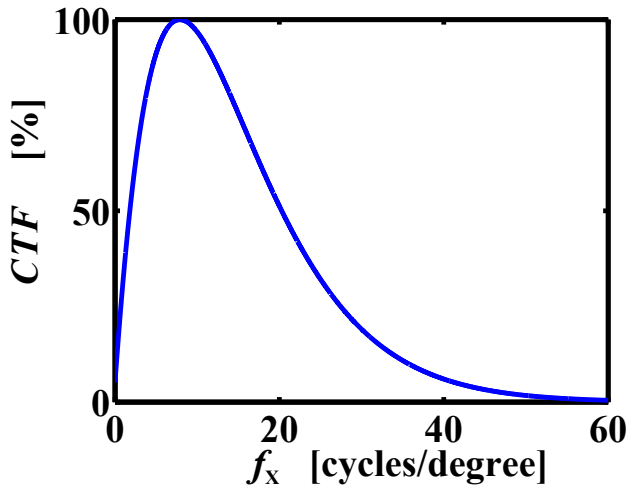


Figure 2.14: *Model of the normalised contrast sensitivity function of the human visual system.*

adaptation to the overall brightness of the room, and variations between individuals such as uncorrected aberrations; these models give an indication of the relative importance of different spatial frequencies in the perceived image. However, to incorporate such a model into a metric of the image quality of an imaging system would require a fixed distance between the subject's pupil and the display. Further simulations presented in this thesis will therefore assume that the contrast at all non-zero spatial frequencies is of equal importance.

2.7 Discussion

Even modest levels of aberration can suppress the MTF below levels that allow restoration of a high fidelity image. For instance, zeros occur in the MTF for $W_{20} > \lambda/2$, leading to irretrievable contrast loss. The introduction of pupil-plane phase-modulation can significantly reduce the MTF variation, and thus increase the defocus-tolerance. It has been shown however that the phase-modulation parameters should be chosen with care; noise amplification and regularisation by the image restoration also reduce the expected image-fidelity. The proposed method allows identification of the phase-modulation parameters that maximise the expected imaging fidelity for a given defocus tolerance. The numerical phase-modulation parameters presented here can be used as initial values for optimisation of specific and rigorously modelled hybrid designs.

The proposed method is insensitive to angular distribution of error. Nevertheless, it is to be expected that an image with angularly varying quality will be perceived differently from an image with an, on-average, uniformly distributed quality. This is pertinent for all antisymmetric masks, particularly the cubic phase-modulation which yields significantly lower errors for horizontal and vertical spatial frequencies as can be

noted from the spoke target images in figure 2.6. A potential improvement to the method presented here could be the incorporation of an angular weighting function to account for perceptual effects of the orientation of errors.

It has been assumed here that the image is Nyquist-sampled. Although many practical imaging systems are under-sampled, the magnitude of errors introduced by aliasing tend to be reduced by the phase-modulation function [102]. The effect of aliasing on the optimum pupil-phase modulation will be discussed in chapter 3.

To estimate the optimal performance, it is calculated here using the optimal deconvolution kernel; that is, the kernel used in image recovery corresponds to the actual out-of-focus PSF. In some applications it may be desirable or necessary to use a single kernel for a wide range of W_{20} and this will introduce image errors that increase with defocus mismatch. In the case of the quartic phase-modulation, these errors are similar to those introduced by regularisation and are well described by the MSSIM metric. However for antisymmetric phase profiles, such as the cubic and generalised-cubic functions, artifacts are introduced that have the form of image replications [47]. An accurate metric of perceived quality of images manifesting this type of artifacts has yet to be defined; however, it is possible that for images for which it is not possible to use the optimal defocus in image recovery, the presence of these image-replication artifacts may yield lower perceived image quality for cubic and generalised-cubic phase functions than for radially-symmetric phase functions.

2.8 Conclusion

Hybrid optical-digital imaging systems offer improved capabilities for reducing the effects of aberrations; however, a new methodology is required to optimise imaging systems for output image quality. Crucial for this is an efficient and accurate method to predict the imaging fidelity of hybrid designs with realistic SNR ratios, thereby accounting for noise amplification and regularisation of the image processing. The method described in this chapter fulfills these requirements and can be used to define optimisation metrics for any optical design for which the PSF can be determined, facilitating integration with existing commercial optical-design software.

Using this technique, commonly studied antisymmetric and radially-symmetric phase functions have been analysed and compared. Numerical simulations show that for a given depth-of-field range, the highest imaging fidelity can be obtained with only two ratios of α and β of the generalised-cubic phase-modulation. Both configurations yield a lower minimum expected imaging error than what can be achieved with the quartic phase-modulation. The imaging fidelity was predicted using a general statistical signal-model, and the results are shown to be compatible with structural similarity evaluations on a set of test-scenes simulated with realistic levels of noise.

Chapter 3 – Simulation of a realistic system

3.1 Introduction

Optical systems consist of a combination of thick lenses or mirrors that at best only approximate the ideal systems considered in chapter 2. Practical constraints such as cost, size, and weight of the lens system limit the the reduction of aberrations. The pupil function will thus not only depend on the object distance but also on the lens-induced aberrations which in general are wavelength and field dependent. Furthermore, vignetting, aperture apodisation, and various losses can affect the modulus of the pupil function. Such aberrations, and the variations thereof with field and wavelength, are design-specific and require numerical methods of evaluation. Nowadays, lens systems are developed with the aid of optical-design software which allows the prediction and minimisation of aberrations before construction. While the reduction of pupil-aberrations is crucial for traditional optical design, the same is not necessarily the case for hybrid optical-digital design. Traditional optical-design software maximises the sharpness of the optical image, and would therefore thwart any intentional pupil-phase modulation, even if such a system permits the reconstruction of a sharper image by digital means. Nevertheless, without modern optical-design software, hybrid optical-digital design would be impractical for realistic lens systems.

This chapter explains how the evaluation method proposed in chapter 2 has been integrated with the off-the-shelf optical-design software package *Zemax*, and how the detection process is simulated more accurately by accounting for the effects of a pixelated detector. Both are studied in detail, revealing some differences with traditional optical-digital design where the pupil modulation is kept minimal. It will be seen that accurate simulation of pupil-modulated optics requires a high-density OTF calculation, both spatially and as a function of wavelength. Furthermore, the orientation of the sensor array in the image plane becomes important for an under-sampled system employing non-axial-symmetric pupil modulation. Its application to the analysis and optimisation of specific hybrid optical-digital designs is postponed to chapter 4.

The current chapter starts with an overview of the considered methods for the calculation of the OTF, central to the metric evaluation. The implementation of the selected method is detailed in section 3.3. The Huygens-Fresnel integral [45, Ch. 4] is calculated in *Matlab* based on a ray-trace of the optical model in *Zemax*. Although this method requires more implementation work, it offers most control over the calculation and therefore accuracy. Somewhat unexpectedly, it also resulted in a considerably more efficient metric calculation. Furthermore, it is shown in section 3.4 how the calculation method can be extended to account for the polarisation and incoherence of light. In section 3.5 the conversion of the Nyquist-sampled OTF to the sampling grid defined by

the detector array is discussed. Many, if not most, digital imaging systems are undersampled which results in aliasing effects; this is particularly true for uncooled infrared imaging systems which are typically sampled at at less than a quarter of the Nyquist rate. Moreover, sensor pixels are sensitive over a finite area, which results in a reduction of the MTF at high spatial frequencies. It will be seen that both effects can be incorporated into the metric calculation.

3.2 Calculation of the optical-transfer function

Various PSF simulations at different wavelengths and field angles are sufficient to characterise a given optical design. These can then be used to simulate the post-processing of a set of test-scenes. The final images can then be evaluated to characterise the performance of the hybrid optical-digital imaging system. When the image restoration process of choice is linear and shift-invariant, the end-to-end point-spread functions of the joint design can be calculated as well. As shown in section 2.5 this can be used directly by some metrics, removing the need for specific test-scenes.

To make automated optimisation practical, the calculation of the quality metric should be efficient, and by consequence also the calculation of the PSF or OTF it is based on. The following sections describe possible calculation or approximation methods of these functions.

3.2.1 Fourier transform of point-spread function

If the PSF is already known, the OTF can be calculated as the Fourier transform of the PSF. For practical situations one would discretise the PSF , and use the two dimensional fast Fourier transform. This algorithm has a time complexity of $O(N_\mathrm{x} \cdot N_\mathrm{y} \log(N_\mathrm{x} + N_\mathrm{y}))$, where N_x and N_y are the dimensions of the sample grid.

To avoid losing information in the sampling process, the PSF should be sampled at the Nyquist frequency of the imaging system. This is possible since each optical system has a finite cutoff spatial-frequency caused by diffraction at the aperture. To avoid information loss due to aliasing, sensor arrays should be sufficiently dense to sample at the Nyquist rate. Nevertheless this is not always the case. Higher density sensor arrays are usually more difficult to produce, and a reduction of the spatial-frequency bandwidth can only be achieved by decreasing the numerical aperture and thus the light-gathering power.

The sampling points on the PSF are usually chosen on a rectangular grid spanning the region where the irradiance has an appreciable value. For diffraction-limited systems the energy of the point-spread function should be strongly centred around its

peak value. For more aberrated systems and in particular systems with intentionally phase-modulated pupils, the PSF will be more extended.

If the phase of the OTF is considered for the quality metric, the choice of the central point of the PSF is important. Shifting the sampling grid by a distance Δx will result in phase changes in the OTF of the form $OTF_{\Delta x}(\mathbf{f}) = OTF_0(\mathbf{f}).\exp(\mathrm{i}\pi\Delta x)$, where $\mathbf{f}$ is the spatial-frequency. This becomes particularly important when combining monochromatic PSFs incoherently to simulate a polychromatic PSF.

3.2.2 Autocorrelation of the exit pupil function

A good approximation of the system's point-spread function is not always directly available. When a description of the optical amplitude of the light wave at the exit pupil is available instead, the magnitude and the phase of the wave at the pupil can be represented as a complex valued function of time and position. For coherent illumination, a reference time can be chosen so that the wave can be described as a complex function of position only.

When the system can be modelled using Fourier optics, the optical-transfer function is the normalised autocorrelation of the pupil function [45, Ch. 6]. This autocorrelation calculation has a time efficiency of $O(N_\mathrm{x}^2 N_\mathrm{y}^2)$ for a $N_\mathrm{x} \times N_\mathrm{y}$ sized grid. In many cases it would be more efficient to calculate the PSF first and from that the OTF, using two fast Fourier transforms. The total time would be of the same order as a fast Fourier transform: $O(N_\mathrm{x} N_\mathrm{y} \log(N_\mathrm{x} + N_\mathrm{y}))$. In the contrary case that only a few values of the OTF are required, the autocorrelation could be more efficient. For k values the time efficiency would be of the order $O(kN_\mathrm{x}N_\mathrm{y})$ for sampled pupil functions.

The size of the required sampling grid might be problematic in case the PSF is spread-out over a large area. This would mean that the pupil function contains significant high frequency components, i.e. it would be highly aberrated. A time efficiency of the order $O(kN_\mathrm{x}N_\mathrm{y})$ could be impractical, and specialised numerical integration techniques for highly oscillatory integrands would be more appropriate [103, 104].

3.2.3 Approximation with ray tracing

For high numerical apertures such as common in microscopy, the polarised nature of the light becomes important. Ray tracing can still be used but with the added complexity that the polarisation direction of the light wave should also be propagated. And since polarisation affects interference, also the calculation of the point-spread function becomes more involved, and one usually has to resort to the polarised point-spread function as described by Richards and Wolf [105].

3.2.4 Calculation of the OTF using Zemax only

Optical design software such as *Code V* or *Zemax* can predict a wide range of performance characteristics before the optics are actually manufactured. By tracing rays through the optical model it is possible to predict the spot size, wavefront shape and many other relevant characteristics as a function of wavelength and field-angle. These tools have become essential in the traditional design process.

The electronic post-processing is an integral part of the hybrid optical-digital imaging system, it is therefore necessary to characterise the system as a whole. Out of the box, optical-design software such as *Code V* or *Zemax* include metrics for this purpose. However, both do offer significant flexibility and extensibility via programming constructs. In principle it is possible to create any metric as long as it can be written in the software's programming language. This has been tested with *Zemax*, first using the merit function editor and later using the built-in macro language `ZPL`.

In *Zemax*, the *merit function editor*, is used to drive the optimisation process. Although it is mainly used to calculate some linear combination of metrics such as spot radius or wavefront deviations, the merit function editor allows the designer to describe more complex functions. Each merit function consists of a list of operands, these can be measurements of aberrations such as the operands `COMA` and `ASTI` respectively measuring the coma and astigmatism of the model. Other operands allow changes on the optical model such as the `CONF`-operand that selects the configuration. A third type of operand applies mathematical functions such as square roots (`SQRT`), and sines (`SINE`) to the output of other operands.

An interesting set of operands to evaluate hybrid imaging systems is the set of MTF-operands. These operands allow the calculation of the MTF value on the sagittal or tangential axis at a chosen spatial-frequency. For all spatial frequencies of interest and within the operational tolerances of the imaging system, these values should be sufficiently large to allow the image to be restored accurately. If additionally these MTF values are approximately constant, a potential defocus-invariant optical design is found.

To test the usefulness of the merit function editor for the evaluation of hybrid imaging systems, a merit function was written to calculate the MTF at four spatial frequencies on the sagittal and tangential axes, for all configurations of an optical system incorporating a cubic phase modulation. The only difference between the configurations is the object distance, set to either 5 m ($W_{20} \approx -2.5\lambda$), 10 m ($W_{20} \approx 0$), or infinity ($W_{20} \approx +2.5\lambda$). The imaging error was calculate for a spatial-frequency-independent signal-to-noise ratio of 100, and the cubic phase-modulation depth α is optimised for a well-corrected f/5 singlet with a large aperture diameter of 1cm at $\lambda = 500\,\text{nm}$.

The MTF is shown for various object distances before and after optimisation in figure 3.1. It can be seen that the optimisation leads to a reduction of the MTF values

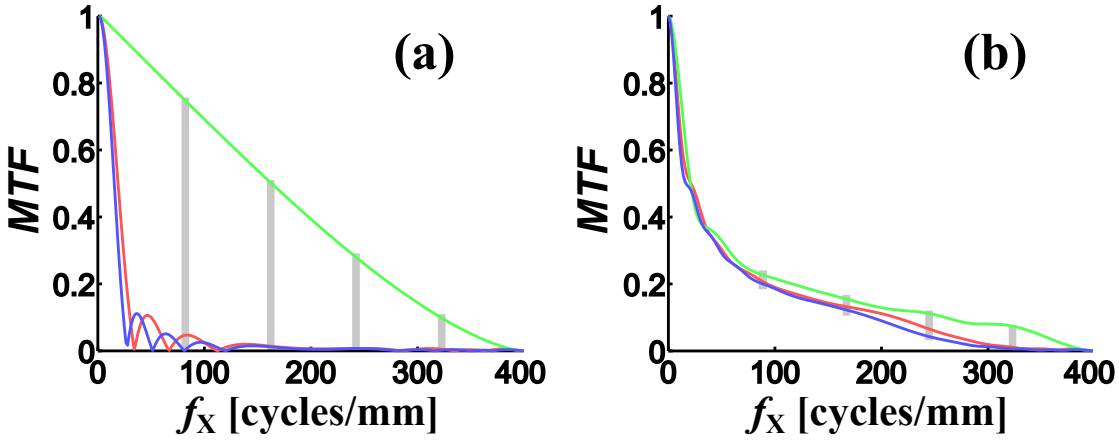


Figure 3.1: *Through focus sagittal modulation transfer functions, (a) before optimisation, (b) after optimisation of the value at the four spatial frequencies indicated with a thick vertical grey line. The coloured solid lines show the MTF for the object distances:* $5m$ *(*$W_{20} \approx -2.5\lambda$*, blue line),* $10m$ *(*$W_{20} \approx 0$*, green line), and* $50m$ *(*$W_{20} \approx 1.5\lambda$*, red line).*

for the in-focus case (green line); however, at the same time it reduces MTF-variation and prevents values too close to zero for the out-of-focus MTFs at 5 m ($W_{20} \approx -2.5\lambda$, blue line), and 50 m ($W_{20} \approx 1.5\lambda$, red line). Note that this method is not restricted to MTF values on the tangential or sagittal axis, by rotating the image plane in various configurations it is possible to sample the two-dimensional MTF at arbitrary spatial frequencies.

Although the merit function editor was clearly not designed for this type of task, it has been shown that it can be used successfully for the design of a defocus-tolerant system. However, the design process is not intuitive and quickly becomes overly complex, e.g. the number of program lines increases with the number of sample points, and the optical configurations have to be used to make small perturbations to the optical model. Furthermore, although the phase contains important information, only the absolute value of the OTF is returned. This is however something that could be solved by using user-defined-operands with the `UDOP` operand.

The *Zemax merit function editor* also features the `ZPLM` operand, this operand allows to invoke *Zemax Programming Language Macros*. In contrast to the merit function editor, it is specifically designed for more complex calculations. It allows a reduction in the complexity of the calculation and makes it more intuitively and less prone to programming errors. The implementation of the MTF-based defocus-tolerance metric described in the previous section was repeated in substantially less time and increasing the MTF sampling density was not considerably less involved.

Although the *Zemax* macro approach has some advantages over using merit function operands only, it is still far from practical. Many of the problems of the previous

approach remain to a certain extent, the phase information is still not available and the optical configurations are still abused to change e.g. the focal distance. Furthermore this macro language is proprietary to *Zemax*, it is unlikely that image restoration algorithms are available in this language. Translation of the algorithms from e.g. *Matlab* or *C* could also be hindered by the limitations of the *Zemax* macro language, in particular the limitation to four arrays will pose significant problems for the implementation of image processing algorithms.

3.2.5 Integration with external software

As demonstrated in the previous section, it is possible to do some performance prediction calculations using only *Zemax*. Both methods presented do have their drawbacks that can be related to the limited expressivity of the languages used to describe the metrics. A third method is therefore considered. *Zemax* also allows various types of machine-compiled extensions; more concretely, inter-process communication is possible using the Microsoft Windows Dynamic Data Exchange (DDE) interface, and the behaviour of the merit function and other parts of the software can be extended with compiled executables or dynamic-link libraries (dll). Practically any metric can thus be devised and implemented in native code, e.g. by compilation of *C* or *C++*. This will certainly result in efficient calculations, however changing and fine-tuning the metric in such a language will require a considerable amount of time invested in taking care of technical details unrelated to the design problem. Higher level languages such as *Matlab* hide many of the technical details from the user, and are often more appropriate for image processing tasks. By consequence, metrics in such a high-level language can be implemented in fewer lines of codes, and in a shorter time due to the reduction in potential errors and the lack of tedious recompilations. To be able to modify the metric with agility for the work presented here, the use of *Matlab* was favoured over low-level languages such as *C* or *C++*. However, when the time needed for optimisation and tolerancing has to be reduced further, regularly used metrics and parameterisations can easily be converted to native code.

The programming language *Matlab* lacks the need for compilation, simplifying the task of the designer significantly. At the same time the programming language is much more suitable for math-intensive task such as image-processing, and the integration of highly optimised matrix-processing libraries such as *LAPACK*[1] and *FFTW*[2] permits effi-

[1] LAPACK, or Linear Algebra PACKage is a software library originally intended for parallelising calculations on super computers. The free availability of its source code has lead to widespread use and subsequent improvements since its initial release in 1992.
`http://www.netlib.org/lapack/`

[2] FFTW is a library with highly optimised implementations of the fast Fourier transform. Initially released in 1997, it also benefitted from accumulated improvements to its freely available source code.
`http://www.fftw.org/`

cient calculations built on the years of code-optimisation invested in these public domain libraries. An additional advantage is that the language is widely used to prototype image-processing algorithms, hence implementations of many algorithms are readily available.

As well as *Zemax*, *Matlab* can be interfaced with compiled machine code. It is thus possible to develop an interface layer in *C++* for communication between *Matlab* and *Zemax*. This is sufficient to automate any simulation that *Zemax* is capable of. However, the results of more specific calculations are not communicated by *Zemax* over the inter-process interface. Crucially, the calculation of e.g. the PSF can only be communicated by prompting *Zemax* to write the data to a file which can be read in from *Matlab*. This process is cumbersome and inefficient, furthermore it is difficult to exercise precise control over the calculation method and the precision of the data format is limited to five significant digits. It was therefore decided to directly interpret the ray-trace information and calculate the PSF in *Matlab*. This avoids the file writing and reading bottleneck, and more importantly, it permits full control over the sample points and calculation method. This is the subject of the following section.

3.3 Scalar calculation of coherent point-spread function of an optical model using ray-tracing

The performance of hybrid optical-digital imaging systems is directly linked to the particular properties of the PSF. Accurate and efficient prediction of the PSF from an optical model is a first essential step in the analysis and simulation of these systems. In this section it is shown how the PSF is calculated from ray traces through an optical model, and in particular how this was implemented using *Matlab* and *Zemax*.

The PSF can in principle be calculated directly by controlling the *Zemax* user interface. There are however several disadvantages to this approach:

1. Importing PSF data is inherently inefficient because it has to be written to a file by *Zemax* and later read and parsed by *Matlab*.

2. Controlling sampling parameters such as the position, size and periodicity of the sampling grid cannot be easily controlled from *Matlab*. This can only be done by rewriting the settings file for every change in settings.

3. The exact calculation done by *Zemax* cannot be verified easily, and the documentation is ambiguous on some points; in example, the exact position of the reference sphere is not well-defined for off-axis PSF calculations. The PSF is central to much of the work presented here, it is therefore necessary to understand its calculation fully.

In contrast, the interface to control ray-tracing itself is efficient and well documented. It was therefore decided to calculate the PSF in *Matlab*, and use *Zemax* only for ray-tracing the optical model. Not only does this give more control over the calculation, it also resulted is significantly more efficient calculations without reducing accuracy.

3.3.1 Scope limitations

In the rest of this document the imaged object will be restricted to a monochromatic unpolarised, and incoherent point source. This allows us to simplify the main discussion considerably while extension to more general scenes is still possible. The following paragraphs discuss these restrictions and their implications briefly.

Ray tracing is in principle a *monochromatic* simulation, polychromatic light is simulated by repeating the process for the range of wavelengths of interest. In the absence of non-linear effects, the PSF can be calculated by integration of the monochromatic PSF intensities over different wavelengths. Only a single wavelength is considered for the main discussion; however, the extension to broadband simulation is discussed in section 3.4.2 with an emphasis on the implication for pupil-engineered optics.

To calculate the intensity pattern on the image plane caused by a point source it is sufficient to know the electrical field distribution at the image plane as a function of time. This is periodic in the case of monochromatic light and the time-variant field distribution can thus be represented by a complex function in space only. The magnitude of the function is the magnitude of the periodic field, and the phase is the relative phase of the field at a specific position.

To limit the scope of this document, only *spatially incoherent* scenes are considered to be imaged. This allows us to make the assumption that light emanating from two distinct points in the scene will not interfere, and therefore the intensity pattern of a point source fully characterises the optical system under investigation if it is known for all possible positions of the point source. Note that the requirement of monochromaticity is in principle incompatible with spatially incoherence; it is therefore assumed that the coherence time is sufficiently short so that a field emanating from two distinguishable points in the scene is uncorrelated, while still sufficiently long to permit a monochromatic approximation of the PSF. The discussion in section 3.4.2, shows that the requirement of a long coherence time can be relaxed by incoherently combining various PSF simulations at different wavelengths.

The *scalar field approximation* will be used in the initial discussion. Polarised ray tracing would give more accurate results, in particular for low f-numbers[3]. For very fast

[3]The f-number is often used instead of numerical aperture (NA) when the latter is much smaller than one. It can be defined as $1/2NA$, and is often noted as f/#, where # is replaced by the value associated with a particular system. Alternative definitions often ignore the refractive index and reduce the f-number to the

systems, the wave's electrical and magnetic components cannot be ignored in the direction perpendicular to the image surface. A more rigourous simulation of polarised light propagation would therefore be more appropriate for systems with numerical aperture close to 1 or larger.

Polarised ray tracing is for a single polarisation roughly a factor of two slower than its unpolarised equivalent. Furthermore, most scenes of interest do not emit perfectly polarised light but emit unpolarised or partially polarised light. For simulating light propagation from these scenes, several ray traces would be required to calculate the combined effect of the various light polarisations. *Zemax* allows the calculation of polarised point-spread functions however only for one polarisation of the source at a time. As will be discussed in section 3.4, this can be used to calculate the PSF more accurately by accounting for polarisation.

3.3.2 Ray tracing in Zemax

Optical design software such as *Zemax* allows us to approximate the propagation of a light wave by ray tracing a bundle of rays through the optical model. Intuitively each ray can be thought of as the centre of a narrow collimated beam of monochromatic light travelling on straight lines between consecutive lens surfaces and refracting at the boundaries between two media. Light emitted from a point source in the scene is simulated by a bundle of rays emerging from this point at different angles. Spot diagrams can be calculated by plotting the ray end-points after refraction through a set of surfaces.

The spot diagrams obtained by calculating the ray-intersections with the image plane give a reasonable first estimate of the PSF, and are useful to estimate the required PSF sampling range (cfr. section 3.3.7). However, an accurate estimate of the point-spread function can only be obtained by accounting for the diffraction. To this end the ray-tracing software must also keep track of the optical path traversed by the rays, or more precisely, the relative phase of its field-components. In *Zemax*, the intersection with the image plane, the direction, optical-path difference, and transmitted intensity can be obtained for each ray traced for a given wavelength, starting point and direction. If required, also the transmitted polarisation components can be obtained for a given initial polarisation. The following sections will show how the rays can be traced to predict the propagation of the electromagnetic waves through the optical model, and how from this the intensity distribution at the image plane can be predicted accurately.

ratio of the focal length and the pupil diameter.

3.3.3 Ray tracing principle and limitations

In most practical optical systems, the wavefronts emanating from a point-source [4] travel along the direction of their normals in homogenous isotropic media. At media boundaries the direction changes following Snell's refraction law. Additionally, the exiting ray has to be in the same plane as the incoming ray and the surface normal. Ray tracing uses this approximation and variations of it for gradient index media to predict the wavefront progress through the optical model.

The ray trace approximation is accurate as long as the wavefront curvature is small and the field amplitude has small relative changes over distances in the order of wavelengths [74, p. 121]. This requirement is in general fulfilled for most optical imaging systems. However there exist two common exceptions: edges such as for the aperture stop, and focal points near media borders, such as at the image plane. The diffraction caused by the edges and the interference at the image plane cannot be accounted for by ray tracing alone. It is common to use the Huygens wavelet integration to calculate the final interference pattern.

3.3.4 Huygens principle

According to the Huygens principle it is possible to calculate the field amplitude and phase as the combined effect of infinitesimal point sources on a surrounding surface. Kirchhoff deduced this more rigorously using Green's theorem on a surface surrounding the image plane and coinciding with the aperture. If both the point source and the image plane are many wavelengths from the aperture, it can be shown that the image-plane intensity depends only on the field distribution at the aperture. The image-plane intensity can be calculated as the integrated effect of infinitesimal point sources with a magnitude equal to the field at the aperture corrected with a so-called obliquity factor. Depending on the choice of approximations, this factor is a function of the position and normal of the integration surface and the angles towards the point source and imaging point where the field has to be calculated. Table 3.1 gives an overview of the different obliquity factors that can be obtained.

where θ_{in} and θ_{out} are the angle between normal to the pupil surface at the integration point on the pupil and the line connecting respectively the source point and the observation point with it.

An integration surface that passes through the aperture stop is often convenient because this allows to integrate over a finite area inside the aperture edge. Outside this area all fields can be assumed to be absorbed by the imaging system construction.

[4] Monochromatic illumination produces a electrical and magnetic field distribution that change periodically in time with a phase and amplitude as a function of position. An electromagnetic wave-front is a surface of constant phase in the optical system under consideration.

$\cos(\theta_{\text{in}})$	Image is distant from the exit pupil (Green)
$\cos(\theta_{\text{out}})$	Object is distant from the entrance pupil (Green)
$\frac{\cos(\theta_{\text{in}})+\cos(\theta_{\text{out}})}{2}$	Huygens
1	Both object and image are distant from the aperture

Table 3.1: *Obliquity factors*

In practice, if there are lens elements between the image plane and the aperture, the geometrical image of the aperture stop is chosen instead. This is not problematic since in general the geometrical optics approximation holds between the aperture stop and the exit pupil.

The integration can be further simplified by assuming that the point sources in the scene are sufficiently far from the aperture stop so that the angle θ_{in} is approximately constant and the obliquity factor only varies with the angle θ_{out}. If furthermore the integration surface is chosen to be spherical and centred near the imaging point, e.g. at the centre of the PSF, then the obliquity factor can be considered to be equal to 1 and thus dropped entirely. Also, the optical-path difference from any point on the integration surface to the centre point is now identical so that the phase delay from the integration surface to the centre point of the reference sphere can be ignored. Unlike the phase which varies with a period proportional to the wavelength, the magnitude can be assumed to be constant near the centre point because the magnitude of the spherical wavelets decreases inversely with distance.

3.3.5 Numerical calculation of Huygens integration

The wavefront evolution cannot be predicted when it reaches the image plane, the increasing curvature of the wavefront cannot be approximated anymore by a planar wavefront perpendicular to a ray. It is nevertheless possible to predict the wavefront evolution to just after the final lens element if this sufficiently far from the image plane[5]. As shown in figures 3.2 and 3.3, the OPD is calculated as the difference between the wavefront and the reference sphere at the exit pupil, the reference sphere is centred at the chief ray intersection with the image surface. As suggested earlier, it is practical to choose the reference sphere surface for the Huygens integration because the phase difference and attenuation for all wavelets is identical between a point on the reference sphere and its

[5]Lenslets in front of each pixel can be considered as part of the sensor, so this is not necessarily a limitation. If all lenslets are identical, its effect on the OTF can be modelled using the pixel transfer function, cfr. section 3.5. Some variation of the transfer function must however be expected with field-angle.

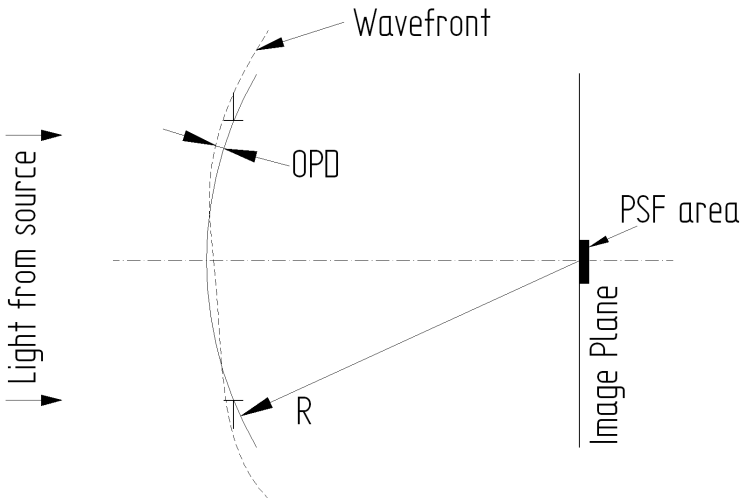


Figure 3.2: *The reference sphere is centred at the PSF centre on the image plane, and has a radius so that it approximately coincides with the exit pupil, which is for this system identical to its aperture stop.*

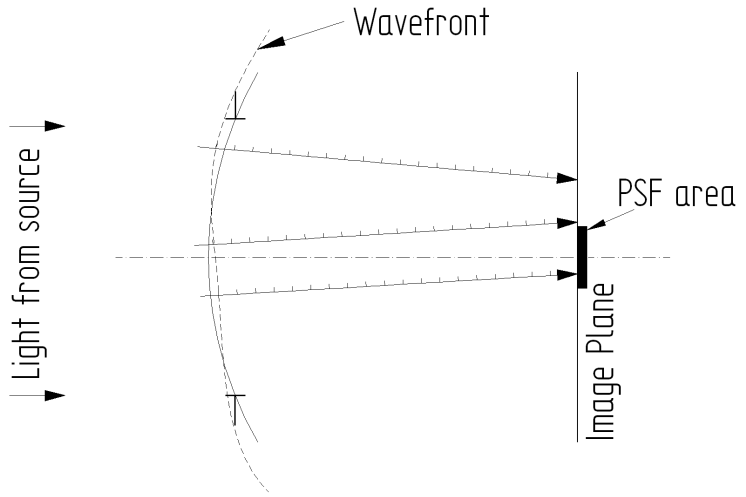


Figure 3.3: *The position of the wavefront is traced with optical path length calculations until it intersects the reference sphere.*

centre point.

In general, the reference sphere radius is large compared to the dimensions of the PSF. Since the field magnitude decreases inversely with the propagation distance, the magnitude of the spherical wavelets is relatively constant for points near the centre of the reference sphere. At the same time, the phase changes across the image with a periodicity that is a function of the angle between the line connecting the Huygens wavelet centre with the point of interest on the image plane and the image plane normal in that point. The field created by each wavelet is thus well approximated by that of a planar wave travelling along a radial of the reference sphere.

If the image surface is planar, and if the spot radius is small compared to the imaging distance, this direction cosine can be approximated by the direction cosine of the

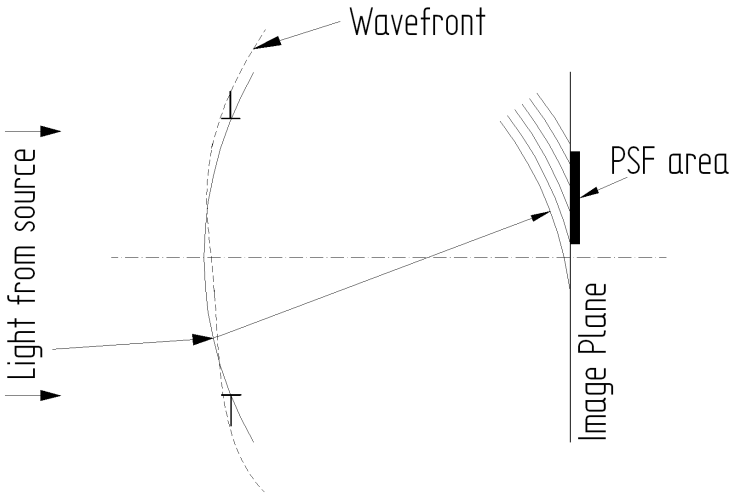


Figure 3.4: *Each point on the reference sphere emits a hypothetical wavelet interfering at the image plane with spherical wavelets emanating from other points on the exit pupil surface. The field near the PSF centre can be approximated by a periodic planar wave if the PSF area is small compared to the radius of the reference sphere.*

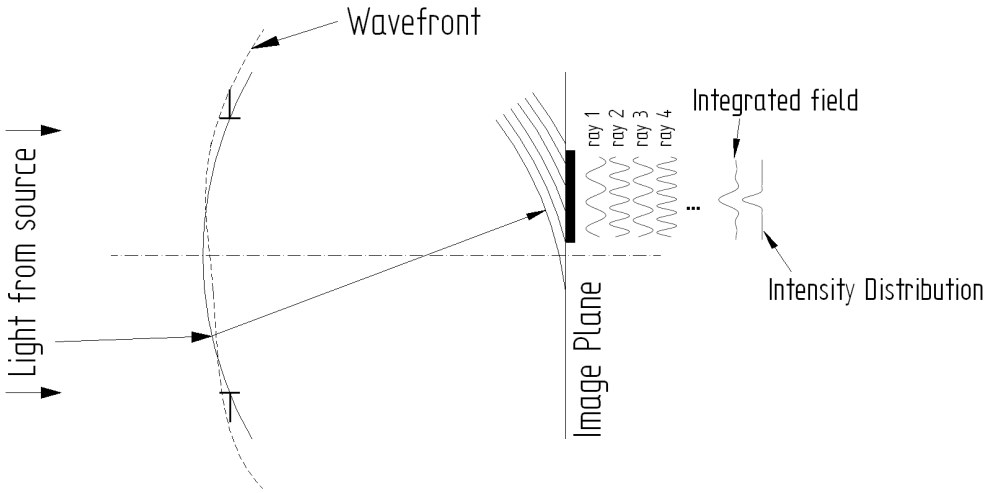


Figure 3.5: *The summation of the periodic fields of the Huygens wavelets allows the calculation of the total field at the image plane. The final PSF intensity is proportional to the squared field.*

ray when it crosses the image surface near the centre of the reference sphere. This is how for the MDO (Multi-Domain Optimisation) project[6] the PSF is calculated, as a periodic electrical field for each wavelet on a grid around the focal point and averaging the effects of all rays(figures 3.4 and 3.5). This assumes that each ray represents an equal fraction of the emitted radiation, which should be approximately true if the entrance pupil is uniformly illuminated and sampled. The main disadvantage of this approach is that the calculation requires a time proportional to the product of the number of rays, N_{rays},

[6]The Multi Domain Optimisation project has been used to optimise hybrid digital-optical thick lens imaging systems. `http://ocpl.ece.arizona.edu/mdo/overview/`

and the number of grid points, N_{gp}. For large aberrations, or pupil-phase-modulated systems, both the number of rays and grid points should be increased for accuracy which can become quickly inefficient.

Note that the *Zemax user's guide*[7] and a treatment of the Huygens wavelet calculation on the *Zemax User's Knowledge base* website[8] describe the Huygens wavelet PSF calculation of *Zemax* also as an integration of periodic fields. However it appears to follow the approach of field integration, leading to significant inefficiencies for the large PSF sizes often encountered in pupil-modulated optics.

3.3.6 Fourier space calculation of the Huygens integral

In the traditional approach set out in the previous subsection, each ray defined a periodic electrical field magnitude around the focal point on the image plane. The PSF is approximated by the intensity distribution of the combined periodical electrical ray fields. The Fourier transform of this summation of the wavelet fields on the image plane can be rewritten as the sum of the Fourier transforms of the periodic effect of each wavelet. The Fourier transform of the electrical field distribution at the image plane focal region can thus be written as a summation of shifted Dirac delta functions multiplied with the electrical field amplitude of the wavelet and with a complex argument equal to the phase. Figure 3.6 shows schematically how the PSF is calculated by doing the integration in the frequency domain.

The position of these Dirac delta functions is only dependent on the direction cosine of the rays and is in general relatively uniform since the rays converge from the exit pupil to typically a smaller area on the sensor. The exit pupil tends to be illuminated relatively uniformly, and the rays can be chosen so that the areas of the wavefront in between rays have approximately the same energy flux.

The Dirac delta function has a non-zero value for only one point, the summation can therefore be done in a time proportional to the number of rays, N_{rays}, instead of the number of the product of the number of rays and the number of grid points, N_{gp}. To calculate the PSF, the sum of the Dirac delta functions has to be inverse Fourier transformed, which can be done in a time in the order of the number of grid points multiplied by its logarithm. The total time taken would be $O(N_{\text{rays}} + N_{\text{gp}} \log(N_{\text{gp}}))$, compared to $O(N_{\text{rays}} N_{\text{gp}})$ for the earlier approach where the periodic electrical field magnitudes are combined in the image space.

Unfortunately the direction cosines do not define a perfectly coinciding grid with

[7] This can be found in the *Zemax* user's guide version of March 6, 2008 p. 126, p. 129 and in particular p. 138.

[8] The description written by Ken Moore in July 2005 as the president of *Zemax Development Corporation* and author of *Zemax* can be found on the url: `http://www.zemax.com/kb/articles/21/4/What-is-a-Point-Spread-Function/Page4.html`

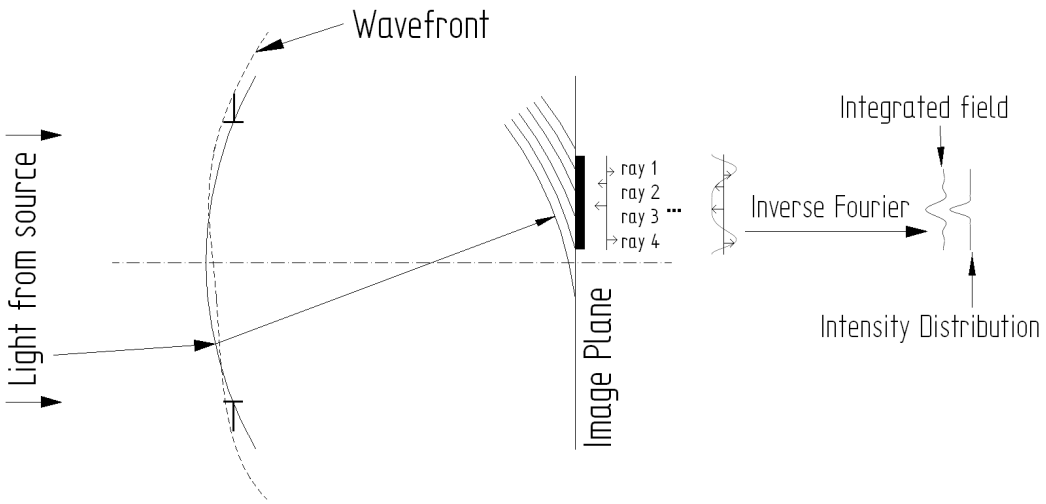


Figure 3.6: *The integration of the periodic fields can also be done in the frequency domain. Conversion is straightforward because the Fourier transform of a periodic field is a Dirac delta pulse. The Dirac delta functions are zero on all but one point, numerical integration is therefore significantly simplified. After integration the field distribution at the image plane can be calculated efficiently with a inverse fast Fourier transform.*

the discrete Fourier sampling positions. The ray tracing can be seen as an irregular sampling of the field distribution's Fourier transform. Small errors can therefore be expected for the described algorithm. For this reason, the implementation of the PSF calculation in this thesis uses interpolation to obtain a regular grid of samples. The interpolation is done linearly on triangles (using Delaunay triangulation) for the phase and amplitude separately, this avoids the bias towards lower amplitudes if the complex-valued pupil function would be interpolated directly. Direct interpolation of the complex phasor would reduce the phasor length and introduce systematic errors. Compared to the Huygens integration in *Zemax* this is a restriction that should be removed for future implementations so that non-uniformly illuminated pupils can also be handled correctly.

It can be seen that the to-be-interpolated function is by definition the pupil function, and is often rather smooth as low order aberrations tend to dominate. When relying on interpolation it is therefore possible to reduce the number of sample points and thus the number of rays to be traced, without trading in accuracy compared to the *Zemax* implementation. While the function is often smooth for most of the area where it has non-zero values, a sharp magnitude change can be expected at the edges of non-apodised apertures. The precise position of this sharp magnitude change defines the cut-off spatial-frequency of the system. An adaptive sampling procedure could therefore increase the accuracy without an efficiency trade-off. The current implementation does not implement adaptive sampling yet, this could give a potential improvement to the algorithm.

3.3.7 Sample grid selection

The PSF is typically calculated on a rectangular grid of sample points on the image surface. A good choice for the size and density of this grid is crucial to limit the calculation time without cropping or under-sampling the PSF. The optimal settings can vary considerably when the optical model is changed, it is therefore necessary to automate the selection of the optimal sampling grid.

All imaging systems have a finite aperture and are therefore band-limited by diffraction. This makes it possible to choose the density of the sampling grid without losing information by the discrete approximation of the continuous intensity distribution on the image plane. According to the Nyquist theorem, the minimal sampling density should be twice the highest spatial-frequency that the imaging system can transmit to the image plane. This is the highest spatial-frequency of the intensity image, or twice the highest spatial-frequency of the field at the image plane. Practically this spatial-frequency can be estimated well by searching for the non-vignetted rays that would produce the highest spatial-frequency on the image plane at the PSF centre. Adding a small margin to the estimated value guarantees sampling above the Nyquist frequency.

While the finite aperture allows us to set an under limit for the sample grid density, it also prevents setting an upper bound to the sample grid size because band-limited signals have in principle an infinite support. The PSF intensity typically decays quickly away from its peak value, but nevertheless it is non-zero over an infinite area.

Fortunately, most point-spread functions have a significant intensity in only a limited region[9]. A pragmatic solution would therefore be to assume that the PSF intensity is non-zero outside a rectangular region enclosing most of the PSF intensity. The size of this region can be chosen as a multiple of the spot radius. However, if this spot radius is small, the size is predominantly defined by diffraction. A secure choice for the size of the calculation region could for instance be calculated as the sum of the spot radius and a multiple of the Airy disc radius.

[9]Exceptions exist where the intensity that is spread out over a large region could still be of practical interest. E.g. a diffraction grating could produce well-separated intensity spots containing significant high spatial-frequency information. Even if high intensity spots are distibuted over a large region, the total area of the image plane where the intensity is non-negligible will still be limited in practical situations. If the intensity is spread out over a large area, its mean and variance must be minimal, and hence also the total variance of the intensity. The total variance of the intensity is proportional to the total variance of the MTF, so the latter will also be close to zero. Hence, point-spread functions covering a large area provide little information for image restoration and will thus be less useful for hybrid imaging system design.

3.3.8 Verification of the PSF calculation

A paraxial lens can be modelled in *Zemax* so that the PSF obtained with the new calculation method can be compared to the analytical expression of the Airy disc[10]. A grid of 256×256 rays is traced through the pupil, and the PSF calculated at 512×512 points separated by $1.25\,\mu m$ as required to avoid under-sampling at a wavelength of $500\,nm$ for an f/5 optical set-up. The root-mean-square difference with the analytical expression of the Airy disc is 0.215% for a PSF with normalised peak intensity. This is practically identical to the value found when the PSF is calculated with the Huygens method in *Zemax*: 0.216%. The root-mean-squared (RMS) difference between the PSF obtained in *Zemax* with the Huygens method and that obtained with the new method proposed in this thesis is negligible at only 0.0005168%.

When the paraxial lens is replaced with a thick singlet lens introducing a wavefront RMS error of 0.2λ, a difference of 0.123% with the analytical expression is found for the new calculation method. However, this difference is small compared to the difference with the PSF calculated with the FFT method of *Zemax*. The PSF obtained with the FFT method has an RMS difference of 0.578% with the Huygens method, and 0.581% with the new method. According to the *Zemax* documentation, several simplifying assumptions are made when using the FFT method, it is therefore less accurate than the Huygens method. Although a small difference was found between the PSF obtained with the new method and that obtained with the Huygens method, it is unclear which of the two is most accurate. The difference between the two methods is deemed to be sufficiently small to justify its use in the proposed *Zemax-Matlab* integration.

Furthermore, it should be noted that on the same standard desktop computer, the Huygens methods requires approximately 550 second to complete, while the new method completes the task in less than 10 seconds. This is a considerable practical advantage because a typical optimisation of an optical model requires several thousands, if not millions of PSF calculations.

3.4 Generalisations of the scalar coherent calculation

Although the scalar monochromatic approximation is often sufficiently accurate, the requirement to operate well for a wide range of wavelengths is common. Furthermore, polarisation influences the point-spread function for systems with a high numerical aperture in image space. In this section it is shown how the method introduced in the previous section can be extended to account for both, polarisation, and broadband illumination. In particular the implications of broadband illumination are considered for

[10]The PSF of a circular aperture is often referred to as the Airy disc, mathematically defined as $I_0 \left(\frac{2J_1(r)}{r} \right)^2$, with r the distance from the centre and I_0 the central intensity.

pupil-engineered optics.

3.4.1 Extension to polarised PSF calculation

So far the propagating light wave has been considered a scalar field wave. This is approximately true for systems with a large f-number or low numerical aperture. For systems with a high numerical aperture, a more accurate calculation can be done when the ray tracing is performed using the polarisation information. Instead of the magnitude and optical-path-difference of each ray, the magnitudes and phases of the planar electrical field along the three spatial dimensions are now obtained. The rays thus sample the Fourier transform of the image plane electrical field distribution's magnitude and phase along three orthogonal spatial directions, the electrical field on the image plane can now be calculated component by component. Assuming linear isotropic media, the computational complexity of this is approximately a factor of three higher for polarised light.

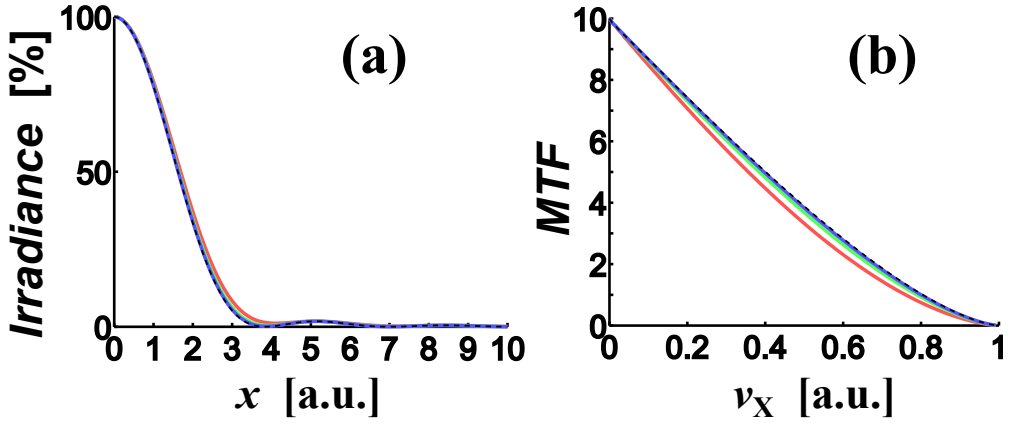


Figure 3.7: *Comparison of the scalar approximation of the PSF (a), and OTF (b), to the full-field calculation of an unpolarised source. The dashed black line corresponds to the scalar approximation. Coloured lines indicate the PSF and MTF of an unpolarised point source, calculated accounting for polarisation [105]. The solid blue, green, and red lines correspond to respectively, an f/2, f/1, and f/0.5 (NA$\approx$ 0.7) system.*

Unpolarised light can be simulated as the incoherent combination of two orthogonal polarisations. The computation time can thus be expected to be significantly longer than that of the scalar approach. Fortunately for most systems the differences are minor between the scalar approach and the specific case of unpolarised illumination. The plot shown in figure 3.7a compares the scalar PSF, obtained using the analytical expression of the Airy disc (dashed black line), to that calculated taking the polarised nature of light into account. Point-spread functions are calculated using the analytical solution described by Richards and Wolf [105], and plotted for various focal ratios: f/2

(solid blue), f/1 (solid green), and f/0.5 or NA≈ 0.7 (solid red). Figure 3.7b shows the corresponding OTFs using the same line types. Both, in figure 3.7a and b, it can be seen that the scalar approximation practically coincides for optical systems as fast as f/1. However, the solid red curve, corresponding to the f/0.5 system, deviates more significantly. Such focal ratios are typically found only in microscopy, however in object space, not in image space.

The additional resources are rarely justified for optical systems with a focal ratio larger than f/1. Although the scalar calculation can be extended readily to account for polarisation, calculations will be done using the more efficient scalar approach for the rest of this thesis.

3.4.2 Polychromatic simulation of broadband light

So far only monochromatic, although incoherent, illumination has been considered. In general, the point-spread function for broadband light is approximated by a linear combination of several monochromatic simulations. Sampling of three to five wavelengths is often considered to be sufficiently accurate. Figures 3.8a and b show for example how in the visible spectrum a diffraction-limited PSF can be approximated well by three monochromatic PSFs. Here the wavelengths $\lambda_F = 486.1327\,\text{nm}$ (blue line), $\lambda_d = 587.5618\,\text{nm}$ (green line), and $\lambda_C = 656.2725\,\text{nm}$ (red line) are used, corresponding to the F, d, and C Fraunhofer lines often used for this purpose. To simulate a uniform spectral density, the intensity of the monochromatic point-spread functions is simply averaged, and overlaid as a black dashed line in the plots. This can be compared to the point-spread function for a dense uniform spectrum shown as a thick grey line. The dense spectrum is simulated by averaging the calculations for 100 wavelengths that are selected uniformly between the λ_F and λ_C-line. Although a difference can be noticed in the width of each monochromatic PSF, the mean PSF closely resembles the actual PSF for a dense spectrum. As can be expected, the same is true for the MTFs shown in figure 3.8c. The black dashed line representing the mean of three monochromatic OTFs almost coincides with the actual MTF shown as a thick grey line. It could even be argued that, in absence of chromatic aberrations, the MTF could be approximated sufficiently well with the central green wavelength, λ_d.

However, as will be shown now, three wavelengths are not necessarily sufficient for pupil-phase-modulated optics. An accurate approximation requires that both the spectral density and the point-spread function variation between wavelength samples is limited. Since the latter extends over a large area when phase-modulation is intentionally introduced, some variation with wavelength must be expected. Even if there are no chromatic aberrations present, the point-spread function will scale proportionally with the wavelength. As shown for a cubic phase-modulation with $\alpha = 3\lambda$ in figure 3.9(a-

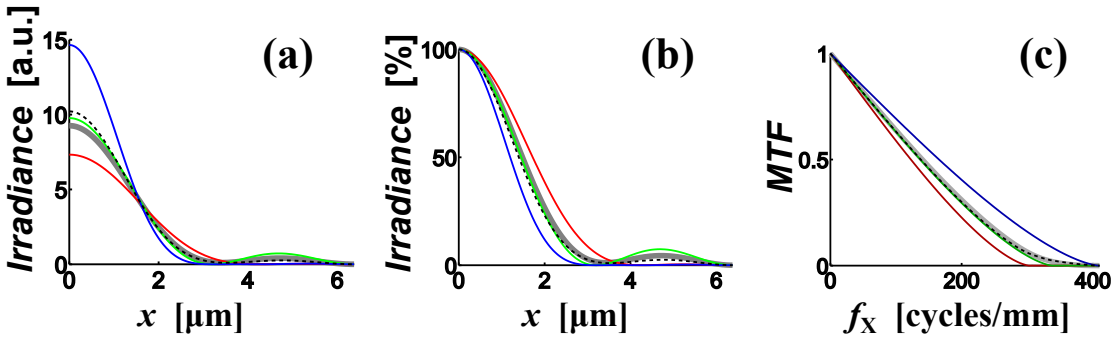


Figure 3.8: *Comparison of an f/5 diffraction-limited system for broadband illumination (thick grey line), and its polychromatic approximation using three wavelengths: $\lambda_F = 486.1327\,nm$ (blue line), $\lambda_d = 587.5618\,nm$ (green line), and $\lambda_C = 656.2725\,nm$ (red line). The mean of the three monochromatic point-spread functions is shown as a dashed black line. Comparisons are shown for: (a) the point-spread functions (each with total irradiance of 100 arbitrary units), (b) the point-spread functions normalised to the on-axis intensity, and (c) the optical-transfer function.*

c), features of the point-spread function that are distant from the Gaussian image point will shift significantly with wavelength. The mismatch becomes more clear when the three monochromatic PSF simulations are combined in the image shown in figure 3.9d. The insufficiency of three wavelengths can be seen when comparing the intensity of the three combined PSF simulations in figure 3.9e, with the PSF for the dense spectrum simulated in figure 3.9f using 100 wavelengths. The densely sampled spectral source is appreciably more blurred than that of the simulation with only three wavelengths. Pupil-phase-modulated optics apparently require a denser sampling of the spectrum for the accurate approximation of the broadband PSF.

The failure to capture the loss of high spatial frequencies using three wavelengths becomes obvious when comparing the MTFs shown in figure 3.10. The same curves as previously shown in figure 3.8c are now plotted for an f/5 optical system with a pupil-phase modulation with amplitude $\alpha = 3\lambda$. The densely sampled MTF shown as a thick grey line can be seen to drop to zero around $f_X \approx 200$ cycles/mm. The mean of the three monochromatic OTFs has a similar behaviour at the lower spatial frequencies; however, while the densely sampled MTF stays near zero for higher spatial frequencies, the approximation with three wavelengths only occasionally reaches zero and has significant values for most of the spatial-frequency spectrum. Note also that since the OTF of this system is complex valued, the combined MTF is lower than the MTF of each of its components.

From the previous discussion it is clear that the typically assumed spectral sampling is not necessarily sufficient for pupil-engineered systems. To be able to understand the requirements for the spectral sampling, a closer analysis of the error of the polychromatic PSF calculation is appropriate.

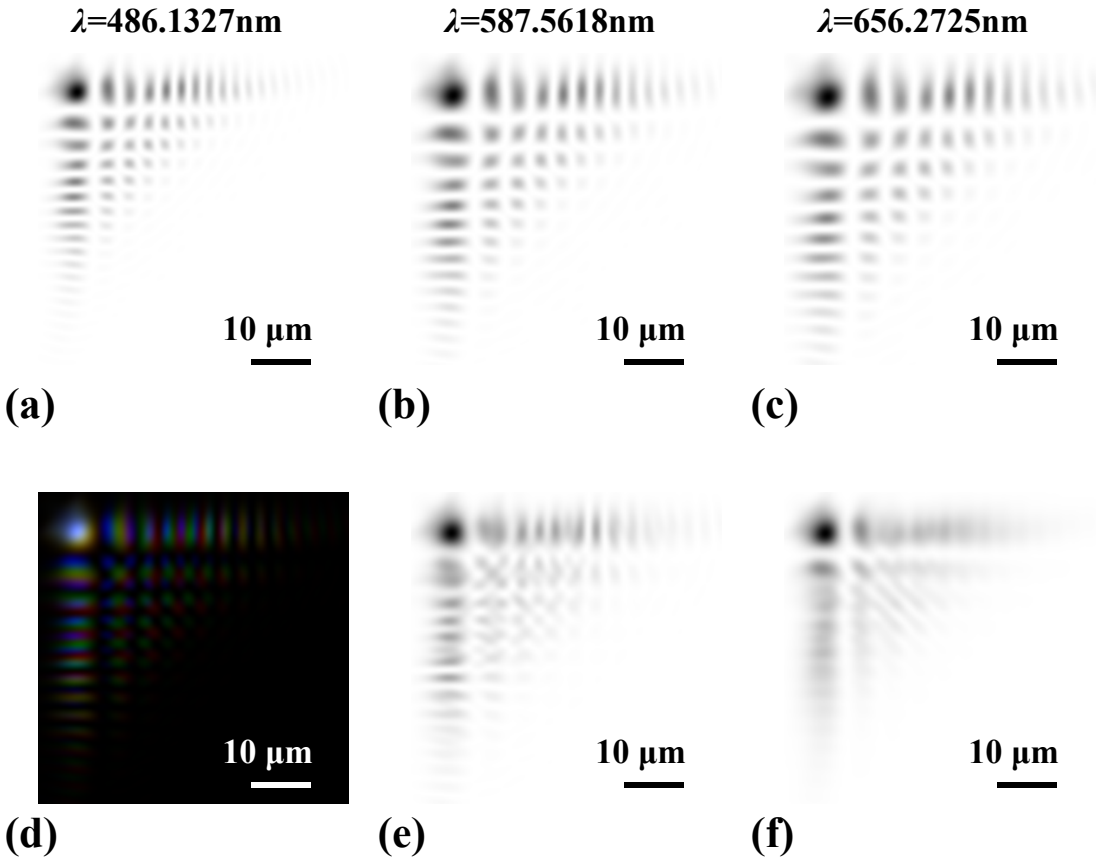


Figure 3.9: *Comparison of an f/5 optical system with cubic pupil-phase modulation of $\alpha = 3\lambda$ for broadband illumination and its polychromatic approximation using three wavelengths: $\lambda_F = 486.1327\,nm$ (a), $\lambda_d = 587.5618\,nm$ (b), and $\lambda_C = 656.2725\,nm$ (c). The combination of the three monochromatic point-spread functions is shown in (d), and the mean irradiance in inverted greyscale in (e). The image shown in (f) is the irradiance for the dense spectrum. All images are on the same scale as indicated by the horizontal bar. With the exception of (e), white indicates zero irradiance, black maximum irradiance. In (e), the irradiance of the three wavelengths is shown in their respective colours with the total intensity normalised for clarity.*

For a spectral density $\Lambda(\lambda)$, the broadband PSF without chromatic aberrations can be written as:

$$\mathrm{PSF}_{\mathrm{bb}}(\mathbf{p}) \quad = \quad \int_{-\infty}^{+\infty} \Lambda(\lambda)\mathrm{PSF}_{\mathrm{d}}\left(\mathbf{p}\frac{\lambda}{\lambda_{\mathrm{d}}}\right)\mathrm{d}\lambda, \tag{3.1}$$

where $\mathrm{PSF}_{\mathrm{d}}(\mathbf{p})$ is the monochromatic PSF for the wavelength λ_{d}, at image plane coordinate $\mathbf{p}$. Stretching the PSF in the image plane is equivalent to shrinking the OTF in the spatial-frequency domain, and since the Fourier transform is a linear operation, the broadband OTF can be written as:

$$\mathrm{OTF}_{\mathrm{bb}}(\boldsymbol{\nu}) \quad = \quad \int_{-\infty}^{+\infty} \Lambda(\lambda)\mathrm{OTF}_{\mathrm{d}}\left(\boldsymbol{\nu}\frac{\lambda_{\mathrm{d}}}{\lambda}\right)\mathrm{d}\lambda, \tag{3.2}$$

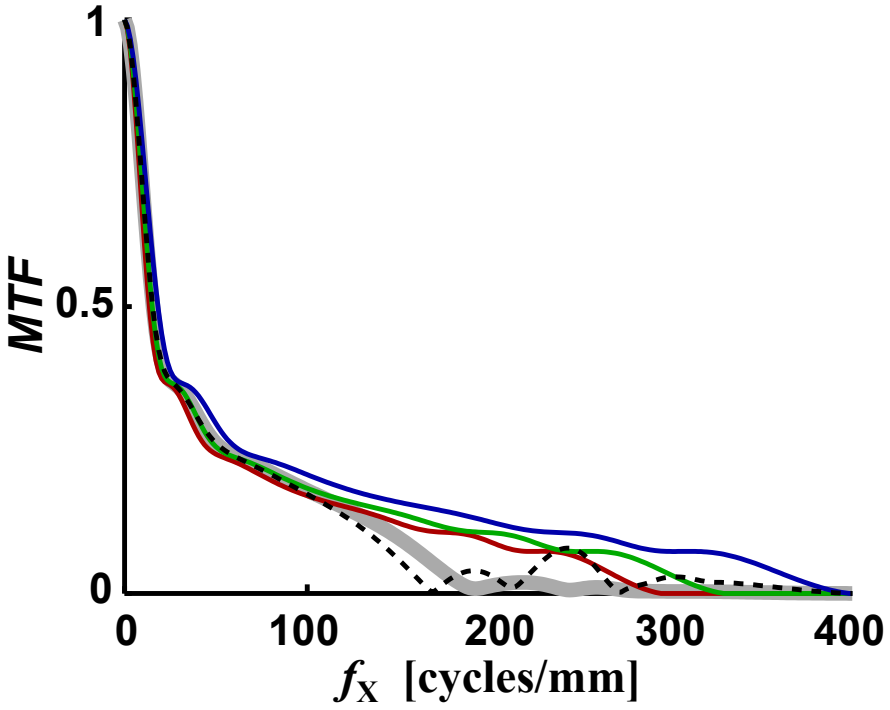


Figure 3.10: *The modulation transfer functions of an f/5 optical system with cubic pupil-phase modulation of $\alpha = 3\lambda$ for broadband illumination. The monochromatic MTFs are shown for three wavelengths: $\lambda_F = 486.1327\,nm$ (blue curve), $\lambda_d = 587.5618\,nm$ (green curve), and $\lambda_C = 656.2725\,nm$ (red curve). The mean of the three monochromatic point-spread functions is shown as a dotted black line, and the MTF corresponding to the dense spectrum between λ_F and λ_D is shown as a thick grey line.*

which for a uniform spectrum of unity spectral density between λ_F and λ_C can be simplified to:

$$\mathrm{OTF}_{\mathrm{bb}}(\boldsymbol{\nu}) \quad = \quad \frac{1}{\lambda_C - \lambda_F} \int_{\lambda_F}^{\lambda_C} \mathrm{OTF}_{\mathrm{d}} \left(\boldsymbol{\nu} \frac{\lambda_d}{\lambda} \right) \mathrm{d}\lambda . \tag{3.3}$$

It can seen from equation (3.3) that in the absence of chromatic aberrations[11], wavelength dependence is equivalent to spatial-frequency dependence of the OTF. An accurate calculation of the broadband OTF, $\mathrm{OTF}_{\mathrm{bb}}(\boldsymbol{\nu})$, requires that rectangle-method integration has a sufficiently small error. If $\mathrm{OTF}_{\mathrm{d}}(\boldsymbol{\nu}\lambda_d/\lambda)$ is twice differentiable, for a sufficiently large number of equally-spaced wavelengths, N, the integration error, ε, of a uniform sampling is bound by the second derivative of the integrand as follows [106]:

$$\varepsilon \quad \leq \quad \frac{N|\lambda_C - \lambda_F|^2}{(N-1)^3 24} \underset{\lambda \in [\lambda_F, \lambda_C]}{\mathrm{argmax}} \left(\frac{\mathrm{d}^2 \mathrm{OTF}_{\mathrm{d}} \left(\boldsymbol{\nu} \frac{\lambda_d}{\lambda} \right)}{\mathrm{d}\lambda^2} \right) . \tag{3.4}$$

[11] Chromatic aberrations are defined here as variations in the pupil function with wavelength.

The OTF of a diffraction-limited system given by equation (2.16) is not twice differentiable at the spatial-frequency cutoff; however, an estimate of the error-bound can be found by approximating the diffraction-limited OTF by the infinite cone $1 - |\|\boldsymbol{\nu}\|\lambda_\mathrm{d}/\lambda|$. Substitution of its second derivative, $2\lambda_\mathrm{d}\|\boldsymbol{\nu}\|/\lambda^3$, in equation (3.4), yields an upper bound for the MTF error:

$$\begin{aligned}
\varepsilon &\leq \frac{N|\lambda_\mathrm{C}-\lambda_\mathrm{F}|^2}{(N-1)^3 24}\underset{\lambda\in[\lambda_\mathrm{F},\lambda_\mathrm{C}]}{\operatorname{argmax}}\left(\frac{2\lambda_\mathrm{d}\|\boldsymbol{\nu}\|}{\lambda^3}\right) && (3.5)\\
&\leq \frac{N|\lambda_\mathrm{C}-\lambda_\mathrm{F}|^2}{(N-1)^3 24}\left(\frac{2\lambda_\mathrm{d}}{\lambda_\mathrm{F}^3}\right) && (3.6)\\
&\lessapprox \frac{N}{(N-1)^3}\frac{|656.2725-486.1327|^2\times 587.5618}{12\times 486.1327^3} && (3.7)\\
&\lessapprox \frac{N}{(N-1)^3}0.0123. && (3.8)
\end{aligned}$$

As could be seen from figure 3.8c for a diffraction-limited system, three wavelengths yields a very close approximation to the actual MTF. This agrees with equation 3.5, which for $N = 3$ limits the OTF error to 1.9%. Four wavelengths will guarantee that the OTF error is never larger than 1%.

For the optical system incorporating a cubic phase-modulation, the second derivative can be calculated using the approximation of the cubic OTF given in [99]:

$$\mathrm{OTF}_\mathrm{cubic}(\|\boldsymbol{\nu}\|) = \frac{1+\mathrm{i}}{\sqrt{96\alpha\|\boldsymbol{\nu}\|}}\mathrm{e}^{4\pi\mathrm{i}\alpha\|\boldsymbol{\nu}\|^3}, \quad \forall\|\boldsymbol{\nu}\|>0\wedge\alpha>0, \tag{3.9}$$

for which the second derivative in equation (3.4) can be calculated as:

$$\frac{\mathrm{d}^2\mathrm{OTF}_\mathrm{cubic}\left(\boldsymbol{\nu}\frac{\lambda_\mathrm{d}}{\lambda}\right)}{\mathrm{d}\lambda^2} = \sqrt{\frac{20736\pi^2\alpha^2\lambda^6\lambda_\mathrm{d}^6\|\boldsymbol{\nu}\|^6+\left(\lambda^6+576\pi^2\alpha^2\lambda_\mathrm{d}^6\|\boldsymbol{\nu}\|^6\right)^2}{768\alpha\lambda^{15}\lambda_\mathrm{d}\|\boldsymbol{\nu}\|}}. \tag{3.10}$$

By verification of the derivatives it can be seen that this expression is maximised when λ is minimised: $\lambda = \lambda_\mathrm{F}$; and for the highest spatial frequencies[12], $\|\boldsymbol{\nu}\| \to 1$. The somewhat simplified expression can be substituted in equation (3.4) to give the following upper

[12]Only values within the interval of validity of equation (3.9) are considered. The OTF approximation of equation 3.9 is not valid for $\boldsymbol{\nu}\to\mathbf{0}$ or $\alpha\to\mathbf{0}$, leading the second derivative to diverge in equation (3.10) for values $\boldsymbol{\nu}\to\mathbf{0}$.

bound for the cubic OTF error:

$$\varepsilon \leq \frac{N|\lambda_C-\lambda_F|^2}{(N-1)^3 24}\sqrt{\frac{20736\pi^2\alpha^2\lambda_F^6\lambda_d^6+\left(\lambda_F^6+576\pi^2\alpha^2\lambda_d^6\right)^2}{768\alpha\lambda_F^{15}\lambda_d}} \tag{3.11}$$

$$\approx\!< \frac{N|\lambda_C-\lambda_F|^2}{(N-1)^3 24}\sqrt{\frac{20736\pi^2\alpha^2\lambda_F^6\lambda_d^6+576^2\pi^4\alpha^4\lambda_d^{12}}{768\alpha\lambda_F^{15}\lambda_d}} \tag{3.12}$$

$$\approx\!< \frac{N}{(N-1)^3}\frac{|\lambda_C-\lambda_F|^2\pi^2}{2}\sqrt{\frac{3\lambda_d^{11}}{\lambda_F^{15}}}\alpha^{3/2}. \tag{3.13}$$

The approximations in the last two steps are made considering that $\lambda_F < \lambda_d$, and that α is sufficiently large. It can be seen that this is a reasonable approximation for $\alpha \geq 1$ wavelength. It can be seen that the error bound increases proportional with the $^3/_2$-root of the phase-modulation amplitude, α, and decreases approximately with the square of the number of wavelengths, N. E.g. if the integration error should be bound to 1%, the number of required wavelengths is:

$$N \approx\!> \sqrt{100\cdot\frac{|\lambda_C-\lambda_F|^2\pi^2}{2}\sqrt{\frac{3\lambda_d^{11}}{\lambda_F^{15}}}\alpha^{3/2}} \tag{3.14}$$

$$\approx\!> \sqrt{100\frac{|656.2725-486.1327|^2\pi^2}{2}\sqrt{\frac{3\times 587.5618^{11}}{486.1327^{15}}}\alpha^{3/2}} \tag{3.15}$$

$$\approx\!> 17\alpha^{3/4}. \tag{3.16}$$

This indicates that 39 wavelengths would be required to maintain the error in the OTF below 1%, considerably more than the typical number of wavelengths used to simulate a traditional optical system. However, it should be emphasised that equation (3.4) expresses an upper bound. In practice fewer wavelengths could be sufficient, in particular if the highest spatial frequencies are expected to be lost due to poor signal-to-noise ratio or under-sampling. It can be seen from equation (3.10) that the second derivative and by consequence the error bound decrease when fewer high spatial frequencies are considered. If the second derivative is calculated for $\nu \leq {}^1/_2$, it can be seen that only six wavelengths are required; while for $\nu \leq {}^1/_4$ two wavelengths are sufficient to keep the integration error below 1%. Nevertheless, it can be concluded that the number of wavelength-samples requires special consideration when dealing with pupil-phase-modulated systems.

3.5 Simulation of detector sampling

In this section, the signal and noise models introduced in the previous chapter are refined for under-sampled imaging systems. With perhaps the exception of those used in

astronomy and microscopy, most digital imaging systems fall into this category. Due to technical and economical reasons there are limits to how small a pixel can be, while at the same time the f-number must be sufficiently low to guarantee an acceptable signal level and field-of-view for a given detector. The finite aperture size limits the highest spatial-frequency that can be found at the image plane: $f_{\text{cutoff}} = {}^{\sqrt{D^2-\lambda^2}}/_{f\,\lambda} \approx {}^{D}/_{f\,\lambda}$, where D is the maximum distance between two points in the aperture, f is the focal length, and λ is the shortest wavelength that can be detected [32]. Even in the absence of detector noise, unambiguous reconstruction[13] is impossible if the pixel centres are further apart than ${}^{1}/_{(2\,f_{\text{cutoff}})}$. This section investigates the imaging error introduced by this ambiguity, as well as the error introduced by pixels with a realistic fill-factor.

In what follows the spatial-frequency domain description is derived for the image of a sensor consisting of a rectangular[14] grid of identical detectors. This description forms the basis of further derivations in this section where the spurious response due to aliasing is separated from the direct response. Although the underlying ideas are introduced in textbooks such as [108], some synthesis would be required from the reader to obtain the complete sampling equation that will be used in this text. In section 3.5.2 it is shown how this description can be employed to predict the imaging error of a sampled imaging system. The derivations are generally applicable to sampled imaging systems, however in section 3.5.3 it is shown that under-sampling has several implications specific to the design of pupil-engineered hybrid imaging systems.

3.5.1 Derivation of the sampling equation

In this section, a concise, however rigourous derivation is given for the analytical expression of the sampled image $i_{\text{samp}}(\hat{x},\hat{y})$ of a continuous irradiance distribution $i(x,y)$ on the sensor plane. For the special case of pixels with a fill-factor of 100%, it will be shown that this can be written in the spatial-frequency domain as:

$$\widehat{I_{\text{samp}}}\left(\hat{X},\hat{Y}\right) \;=\; \operatorname{sinc}\left(\hat{X}\right)\operatorname{sinc}\left(\hat{Y}\right) I\left(\hat{X},\hat{Y}\right) \otimes \text{Ш}\left(\hat{X},\hat{Y}\right), \tag{3.17}$$

with $\operatorname{sinc}(x) = {}^{\sin(\pi x)}/_{\pi x}$, and where $\widehat{I_{\text{samp}}}\left(\hat{X},\hat{Y}\right)$ is the Fourier transform of the sampled irradiance, defined only for discrete spatial frequencies $\hat{X}$ and $\hat{Y}$. The Fourier transform of the image plane irradiance is represented by the continuous function $I\left(\hat{X},\hat{Y}\right)$. The symbol $\otimes$ represents the convolution operator, and $\text{Ш}\left(\hat{X},\hat{Y}\right)$ is the two-dimensional Dirac

[13] The reconstruction of the bandlimited optical image can be done by sinc-interpolation of the signal recorded by point-detectors. More realistic detectors would also require the inversion of the pixel transfer function, which is typically void of nulls in the pass-band.

[14] Colour imaging using a Bayer pattern filter is somewhat complicated by the fact that the green pixels do not form a complete rectangular grid. The irradiance at the intermediate positions is usually estimated by interpolation. This and other types of non-rectangular arrays such as the log-polar layout developed by Wodnicki et al. [107], are considered to be outside the scope of this document.

comb function[15]. Distances are expressed in units of sample distance, hence spatial frequencies are defined for integer fractions.

As a first step, the sampling equation is derived for a rectangular grid of point detectors, i.e. pixels that are only sensitive at coordinates $(x,y)|(\frac{x-o_x}{s_x},\frac{y-o_y}{s_y}) = (\hat{x},\hat{y}) \in \mathbb{Z}^2$, where $(s_x,s_y) \in \mathbb{R}^2$ and $(o_x,o_y) \in \mathbb{R}^2$ are respectively the detector spacing and offset in the x-y-coordinate system. To avoid complicating the notation, coordinates are measured in units of pixel spacing so that $s_x = s_y = 1$, and therefore $\hat{x} = x - o_x$, and $\hat{y} = y - o_y$. For the case of point detectors, i.e. zero fill-factor pixels, the image samples are proportional to the irradiance at the corresponding position:

$$i_{\text{point}}(\hat{x},\hat{y},o_x,o_y) \quad = \quad s_{\text{point}} \cdot i(\hat{x}+o_x,\hat{y}+o_y), \quad \forall(\hat{x},\hat{y}) \in \mathbb{Z}^2 \tag{3.18}$$

where s_{point} is the sensitivity of the point detector, indicating the efficiency with which irradiance is converted to a signal.

Unlike point detectors, real pixels are sensitive over a finite area. The signal of a real pixel can be calculated by integrating equation (3.18) over the grid-offsets o_x and o_y. To account for sensitivity variations across the pixel surface, s_{point} can be replaced by the spatially-variant pixel response function $h_{\text{pix}}(x-\hat{x},y-\hat{y})$. For identical pixels on a rectangular grid, the sampling value at $(\hat{x},\hat{y})$ can now be written as the integral:

$$\begin{aligned} i_{\text{samp}}(\hat{x},\hat{y}) &= \iint_{-\infty}^{+\infty} h_{\text{pix}}(o_x,o_y) \cdot i(\hat{x}+o_x,\hat{y}+o_y)\mathrm{d}o_x\mathrm{d}o_y \\ &= h_{\text{pix}}(-\hat{x},-\hat{y}) \otimes i(\hat{x},\hat{y}), \end{aligned} \tag{3.19}$$

where $\otimes$ denotes the two-dimensional convolution of two continuous functions.

Note that the obtained function $i_{\text{samp}}(\hat{x},\hat{y})$ is not discrete, it is defined and not necessarily zero for $(\hat{x},\hat{y}) \notin \mathbb{Z}^2$. This is not an issue as long as further analysis is restricted to integer coordinates. However, continuous operations such as the Fourier transform are not meaningful unless $i_{\text{samp}}(\hat{x},\hat{y})$ is defined for arguments $(\hat{x},\hat{y}) \in \mathbb{R}^2$. A direct correspondence with the discrete Fourier transform exist when sinc-interpolation is used, to define the signal it is therefore convenient to use this for the rest of this discussion. Mathematically the sinc-interpolation can be written as a multiplication with a two-dimensional Dirac comb function, followed by a convolution with $\text{sinc}(\hat{x})\text{sinc}(\hat{y})$:

$$\begin{aligned} \overline{i_{\text{samp}}}(\hat{x},\hat{y}) &= i_{\text{samp}}(\hat{x},\hat{y}) \cdot \text{Ш}(\hat{x},\hat{y}) \otimes \text{sinc}(\hat{x})\text{sinc}(\hat{y}) \\ &= \left[h_{\text{pix}}(-\hat{x},-\hat{y}) \otimes i(\hat{x},\hat{y})\right] \cdot \text{Ш}(\hat{x},\hat{y}) \otimes \text{sinc}(\hat{x})\text{sinc}(\hat{y}). \end{aligned} \tag{3.20}$$

[15]The two-dimensional Dirac comb function $\text{Ш}(x,y)$ is defined as $\text{Ш}(x) \cdot \text{Ш}(y)$, with $\text{Ш}(x) = \sum_{n=-\infty}^{+\infty} \delta(x-n)$, and $\delta(x)$ the Dirac delta, equal to 0, $\forall x \neq 0$, and integrating to 1.

Note that convolution with the sinc-function leaves the values at integer indexes unaltered, in other words, it adheres to the definition of an interpolation.

The Fourier transform $\overline{I_{\text{samp}}}(\hat{X},\hat{Y})$ of the discretised sampling is:

$$\begin{aligned}\overline{I_{\text{samp}}}(\hat{X},\hat{Y}) &= I_{\text{samp}}(\hat{X},\hat{Y}) \otimes \text{Ш}(\hat{X},\hat{Y}) \cdot \sqcap(\hat{X},\hat{Y}) \\ &= \left[H_{\text{pix}}(\hat{X},\hat{Y})^* \cdot I(\hat{X},\hat{Y})\right] \otimes \text{Ш}(\hat{X},\hat{Y}) \cdot \sqcap(\hat{X},\hat{Y}),\end{aligned} \tag{3.21}$$

where the pixel transfer function $H_{\text{pix}}(\hat{X},\hat{Y})$, and $I(\hat{X},\hat{Y})$ are the Fourier transform of respectively the pixel sensitivity function and the image-plane intensity, and the symbol * denotes the complex conjugate. The two-dimensional rectangle function[16] $\sqcap(\hat{X},\hat{Y})$ is the Fourier transform of $\text{sinc}(\hat{x})\text{sinc}(\hat{y})$.

Figure 3.11 demonstrates the formation of the sampled image in the spatial-frequency domain for a point source. Figure 3.11a shows the Fourier transform of the optical image, the irradiance distribution at the image plane. As can be seen from the circular contour line, the diffraction-limited MTF reaches zero for $\|\boldsymbol{\nu}\| = 1$. When accounting for a fill-factor of 100%, it can be seen from figure 3.11b that the transfer function decreases more rapidly and reaches zero at four straight contour lines forming a square. At these points the pixel transfer function, $H_{\text{pix}}(\hat{X},\hat{Y})$, passes through zero. As will be discussed in the following paragraphs, this is the case for rectangular pixels with uniform sensitivity. Figure 3.11c shows the effect of the convolution with $\text{Ш}(\hat{X},\hat{Y})$ in equation (3.21). This convolution adds shifted replicas of $\mathscr{F}\left\{i_{\text{samp}}(\hat{x},\hat{y})\right\}$ at every spatial-frequency unit when distances are measured in pixels, hence at $1/3^{th}$ for three times under-sampling as shown in figure 3.11c. The rectangle function selects the single period of the added-together replicas shown in figure 3.11d, and sets the remaining periods to zero. The multiplication with the rectangle function is of limited practical use in the analysis; it is therefore custom to remove it from the equation, equivalent to assuming Dirac-delta-interpolation of the samples instead of sinc-interpolation. Equations (3.20) and (3.21) can then be written as:

$$\widehat{i_{\text{samp}}}(\hat{x},\hat{y}) = [h_{\text{pix}}(-\hat{x},-\hat{y}) \otimes i(\hat{x},\hat{y})] \cdot \text{Ш}(\hat{x},\hat{y}) \tag{3.22}$$

$$\widehat{I_{\text{samp}}}(\hat{X},\hat{Y}) = \left[H_{\text{pix}}(\hat{X},\hat{Y})^* \cdot I(\hat{X},\hat{Y})\right] \otimes \text{Ш}(\hat{X},\hat{Y}). \tag{3.23}$$

Note that a discrete Fourier transform of $N_{\text{x}} \times N_{\text{y}}$ samples would return the follo-

[16] The two-dimensional rectangle function is defined as

$$\begin{aligned}\sqcap(x,y) &= 1, \quad \forall x,y \in \mathbb{R}, |x| < 1/2 \wedge |y| < 1/2 \\ &= 1/2, \quad \forall x,y \in \mathbb{R}, |x| = 1/2 \wedge |y| = 1/2 \\ &= 0 \quad \text{otherwise},\end{aligned}$$

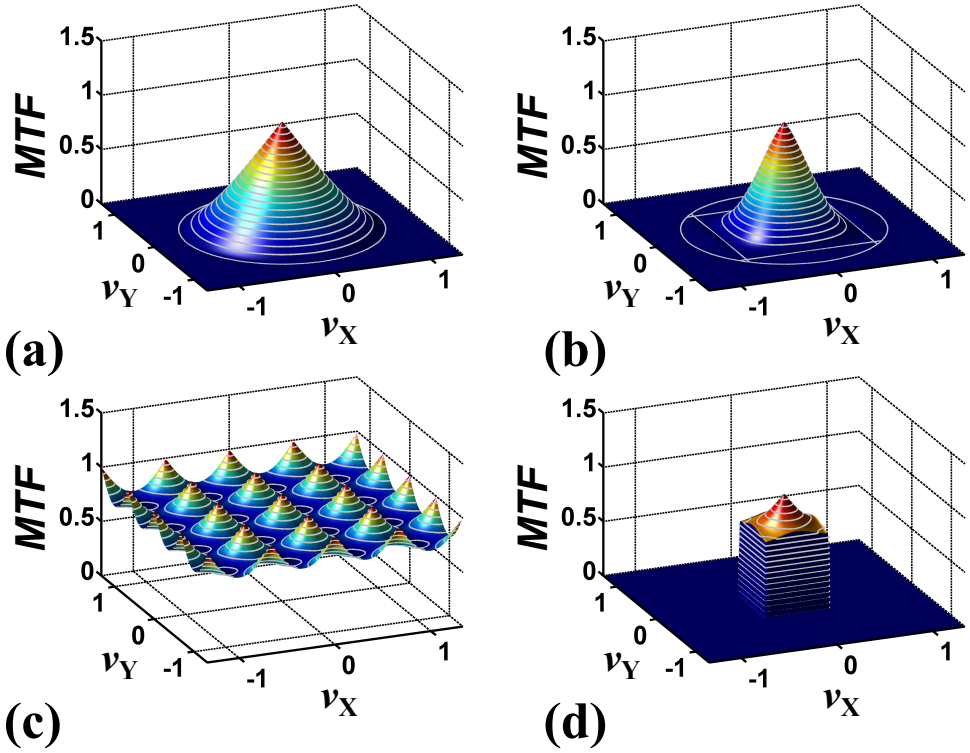


Figure 3.11: *The formation of the sampled transfer function following the operations of the sampling equation (3.21) for a three times under-sampled imaging system. The diffraction-limited transfer function shown: (a) for a perfect measurement of the optical image, (b) accounting for a pixel with fill-factor of 100%, (c) including the replications caused by the convolution with a Dirac comb, (d) selection of the in-band spatial frequencies, up to half the Nyquist frequency of the detector. Spatial frequencies are normalised to the optical cut-off.*

wing values for integer indexes k and l:

$$\begin{aligned} f_{k,l} &= N_\mathrm{x} N_\mathrm{y} \cdot \widehat{I_\mathrm{samp}}\left({}^{k}/_{N_\mathrm{x}}, {}^{l}/_{N_\mathrm{y}}\right) \\ &= N_\mathrm{x} N_\mathrm{y} \cdot \overline{I_\mathrm{samp}}\left({}^{k}/_{N_\mathrm{x}}, {}^{l}/_{N_\mathrm{y}}\right) \quad \forall k,l \in \mathbb{Z}, |k| < N_\mathrm{x}/2 \wedge |l| < N_\mathrm{y}/2, \end{aligned} \tag{3.24}$$

albeit in a more convenient order avoiding negative indexes.

The pixel sensitivity function is often modelled with the two-dimensional rectangle function $h_\mathrm{pix}(x,y) = \sqcap(x/\phi_x, y/\phi_y)$, where ϕ_x and ϕ_y are the fraction of the sensitive area in both dimensions, hence the product of both is the fill-factor of the pixel. The Fourier transform of the pixel sensitivity is:

$$H_\mathrm{pix}(\hat{X},\hat{Y}) = \mathscr{F}\left\{\sqcap\left({}^{x}/_{\phi_x}, {}^{y}/_{\phi_y}\right)\right\}_{(x,y)\to(\hat{X},\hat{Y})} = \mathrm{sinc}(\hat{X}\phi_x)\mathrm{sinc}(\hat{Y}\phi_y). \tag{3.25}$$

The extended sampling equation for the rectangular pixel sensitivity can be written in

the spatial-frequency domain as:

$$\widehat{I_{\text{samp}}}(\hat{X},\hat{Y}) \;=\; \left[\text{sinc}(\hat{X}\phi_x)\text{sinc}(\hat{Y}\phi_y)\cdot I(\hat{X},\hat{Y})\right] \otimes \text{Ш}(\hat{X},\hat{Y}), \tag{3.26}$$

which when converted back to the initial length units becomes:

$$\widehat{I_{\text{samp}}}(s_x\hat{X},s_y\hat{Y}) \;=\; \left[\text{sinc}(s_x\hat{X}\phi_x)\text{sinc}(s_y\hat{Y}\phi_y)\cdot I(s_x\hat{X},s_y\hat{Y})\right] \otimes \text{Ш}(s_x\hat{X},s_y\hat{Y}), \tag{3.27}$$

with the reintroduction of the detector pixel separations s_x and s_y.

It can now be seen why zeros are found at $|\nu_{\text{X}}| = {}^2\!/_3{}^{\text{th}}$ and $|\nu_{\text{Y}}| = {}^2\!/_3{}^{\text{th}}$ in figure 3.11b, at these lines $\text{sinc}(\hat{X})$ or $\text{sinc}(\hat{Y})$ cross zero in equation (3.27) for a three times under-sampled system with pixels of a 100% fill-factor[17]. Notice however that typically only spatial frequencies up to half the Nyquist frequency of the detector are considered, in this case the MTF reduces only to ${}^2\!/_\pi \approx 64\%$ of the value of the optical-transfer function. Even more modest reductions will be found for pixels of a fill-factor less than 100%.

3.5.2 The imaging error of a sampled imaging system

Equation (3.23) shows that both the pixel transfer function, $H_{\text{pix}}(\hat{X},\hat{Y})$, and the replications due to the convolution with $\text{Ш}(s_x\hat{X},s_y\hat{Y})$ perturb the recording. The pixel-transfer function can be accounted for in the treatment of chapter 2 by replacing the optical-transfer function H_{ab} in equation (2.9) with $H_{\text{pix,ab}}(\hat{X},\hat{Y}) = H_{\text{pix}}(\hat{X},\hat{Y})^* \cdot H_{\text{ab}}(\hat{X},\hat{Y})$. Some complications arise when extending the method to under-sampled systems; however, in the following paragraphs it will be seen that the method of chapter 2 can also account for under-sampled systems.

For under-sampled imaging systems also the convolution with $\text{Ш}(s_x\hat{X},s_y\hat{Y})$ has to be accounted for. In such systems aliasing will occur, i.e. higher spatial-frequency components will be present in the recording of components at lower spatial frequencies. The recorded spatial-frequency component will be a linear combination of the corresponding component in the optical image, and one or more components of different spatial frequencies:

$$\widehat{H_{\text{samp}}}(\hat{X},\hat{Y}) \;=\; H_{\text{pix,ab}}(\hat{X},\hat{Y}) \otimes \text{Ш}(\hat{X},\hat{Y}) \tag{3.28}$$

$$\;=\; \sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l), \tag{3.29}$$

where length is measured in units of the pixel separation. The system transfer function,

[17] The Nyquist frequency of a three times under-sampled imaging system is ${}^1\!/_3{}^{\text{th}}$ of the optical cut-off spatial-frequency, hence the detector pixel separations are $s_x = s_y = {}^3\!/_2$ when using spatial frequencies normalised to the optical cut-off.

$\widehat{H_{\text{samp}}}(\hat{X},\hat{Y})$, can be separated into a direct component and a spurious component as follows:

$$\widehat{H_{\text{samp}}}(\hat{X},\hat{Y}) = H_{\text{direct}}(\hat{X},\hat{Y}) + H_{\text{spurious}}(\hat{X},\hat{Y}) \tag{3.30}$$

$$\widehat{H_{\text{direct}}}(\hat{X},\hat{Y}) = H_{\text{pix,ab}}(\hat{X},\hat{Y}) \tag{3.31}$$

$$\widehat{H_{\text{spurious}}}(\hat{X},\hat{Y}) = \sum_{(k,l)\in\mathbb{Z}^2\setminus\{(0,0)\}} H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l). \tag{3.32}$$

This separation is shown in figure 3.12 for a three times under-sampled imaging system. Figure 3.12a shows the Fourier transform of the sampled image of a point source as a function of normalised spatial frequencies ν_X and ν_Y. The direct and spurious parts are shown respectively in figures 3.12b and c. Typically only the spatial frequencies up to half the Nyquist frequency of the detector are of interest[18], this central region of the surface plots in figures 3.12(b-c) is shown enlarged in figures 3.12(d-f) respectively. All the recorded information is present in the direct component; however, it can be seen from figures 3.12e and f that a significant part of the recorded signal consist of spurious components. Typically the low spatial frequencies are less affected, hence the spurious response can be considered to reduce the resolution further than the sensor resolution.

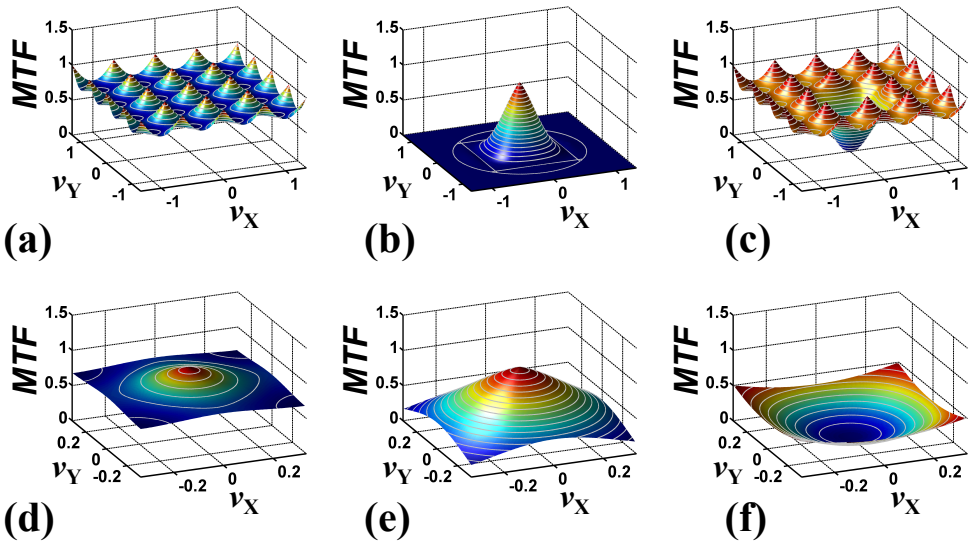


Figure 3.12: *(a) The modulus of the Fourier transform of the recorded image of a point source projected on the centre of a detector pixel. (b) The direct component, (c) the spurious components. (d-f) enlarged central section of the above plots.*

Note that only the direct transfer function is shift-invariant. Figure 3.12 shows the Fourier transforms of the sampled image of a point-source that is projected to the centre of a pixel. If the point source is projected in between two pixels, the plots shown in fi-

[18]Higher spatial frequencies are indirectly available from the replicas. Although this is not considered further in this text, it can allow the reconstruction of super-resolution images with a resolution as high as the optical cut-off spatial-frequency [109].

gure 3.13 are obtained. Notice that only the direct response shown in figures 3.13b and e remains the same, the different interference between Fourier-components creates an asymmetry between ν_X and ν_Y. It is clear that the influence of the spurious components on the recorded image cannot be obtained from a single impulse response. However, in continuation it will be shown that a signal model combined with the direct response, both readily available in simulations, is sufficient to predict the imaging error caused by the spurious components.

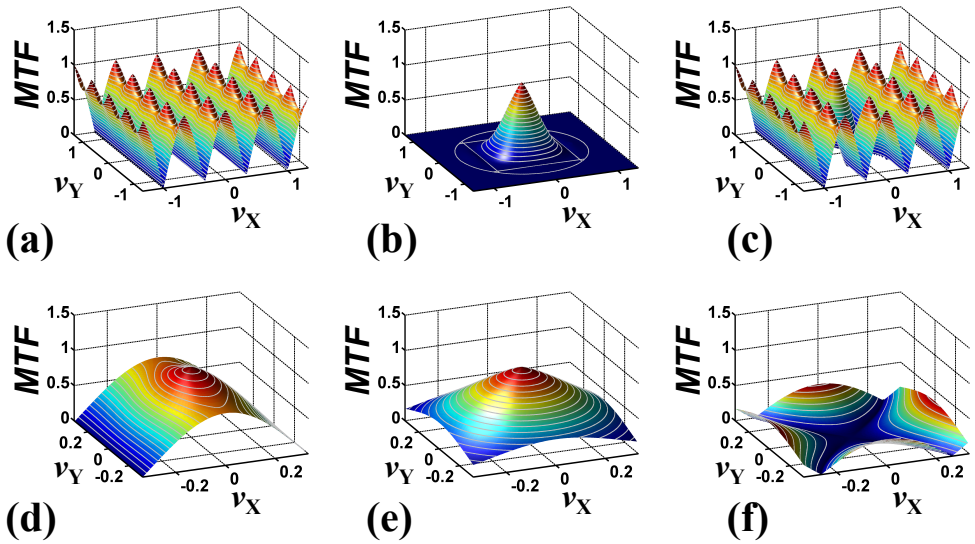


Figure 3.13: *(a) The modulus of the Fourier transform of the recorded image of a point source projected between the centres of two detector pixels. (b) The direct component, (c) the spurious components. (d-f) enlarged central section of the above plots.*

Unlike typical sensor noise, the spurious components introduce an error that can be expected to be highly variant with spatial-frequency. The expectation value of this error is given by:

$$\overline{\varepsilon_{\text{spurious}}}(\hat{X},\hat{Y}) = \mathbb{E}\left\{\sum_{(k,l)\in\mathbb{Z}^2\setminus\{(0,0)\}} H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l)\cdot S(\hat{X}-k,\hat{Y}-l)\right\} \quad (3.33)$$

$$= \sum_{(k,l)\in\mathbb{Z}^2\setminus\{(0,0)\}} \mathbb{E}\left\{H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l)\cdot S(\hat{X}-k,\hat{Y}-l)\right\} \quad (3.34)$$

$$= \sum_{(k,l)\in\mathbb{Z}^2\setminus\{(0,0)\}} H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l)\cdot \mathbb{E}\left\{S(\hat{X}-k,\hat{Y}-l)\right\}. \quad (3.35)$$

The argument of a non-zero spatial-frequency in the scene, $S(\hat{X},\hat{Y})$, must be uniformly distributed when considering all scenes to be equally likely. This can be seen by separating this set in classes of scenes that apart from a lateral shift are identical. As inside each class all shifts are equally likely, the argument must be uniformly distributed because it is proportional with the shift distance. Furthermore, in each class the modulus is identical for a specific spatial-frequency, hence the expectation value of $\mathbb{E}\left\{S(\hat{X},\hat{Y})\right\}$ must be zero

for each class, and by consequence also for the set of all scenes.

The variance of the spurious error is not zero as can be seen from the following derivation:

$$\begin{aligned}
\overline{\varepsilon^2_{\text{spurious}}}(\hat{X},\hat{Y}) &= \mathbb{E}\left\{\left|\sum_{(k,l)} H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l)\cdot S(\hat{X}-k,\hat{Y}-l)\right|^2\right\} && (3.36)\\
&= \mathbb{E}\left\{\sum_{(k,l)}\left|H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l)\cdot S(\hat{X}-k,\hat{Y}-l)\right|^2\right. && (3.37)\\
&\qquad \left. -2\sum_{\substack{(k,l),\\(m,n)}} \Re\left\{\begin{matrix} H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l)^* H_{\text{pix,ab}}(\hat{X}-m,\hat{Y}-n)\cdot \\ S(\hat{X}-k,\hat{Y}-l)^* S(\hat{X}-m,\hat{Y}-n)\end{matrix}\right\}\right\}\\
&= \sum_{(k,l)}\left|H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l)\right|^2\cdot\mathbb{E}\left\{\left|S(\hat{X}-k,\hat{Y}-l)\right|^2\right\} && (3.38)\\
&\qquad -2\sum_{\substack{(k,l),\\(m,n)}} \Re\left\{\begin{matrix} H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l)^* H_{\text{pix,ab}}(\hat{X}-m,\hat{Y}-n)\cdot \\ \mathbb{E}\left\{S(\hat{X}-k,\hat{Y}-l)^* S(\hat{X}-m,\hat{Y}-n)\right\}\end{matrix}\right\},
\end{aligned}$$

with $(k,l)\in\mathbb{Z}^2\setminus\{(0,0)\}$, $(m,n)\in\mathbb{Z}^2\setminus\{(0,0),(k,l)\}$, and where $\Re\{\cdot\}$ refers to the real part of its argument. The expectation value in the second summation in equation (3.39) is zero when the argument of the two factors is independent and uniformly distributed. It was already seen that the argument of each spatial-frequency component separately must be uniformly distributed by considering a class of scenes that only differ in lateral shift. The same is true for the difference between the argument of two distinct components, because a shift in the right direction will change both arguments proportional with the shift distance, yet with a different factor. The expectation value of the second summation in equation (3.39) is therefore zero, and the variance of the spurious error can thus be written as a linear combination of the power spectrum of the scene irradiance at one or more spatial frequencies:

$$\overline{\varepsilon^2_{\text{spurious}}}(\hat{X},\hat{Y}) = \sum_{(k,l)\in\mathbb{Z}^2\setminus\{(0,0)\}}\left|H_{\text{pix,ab}}(\hat{X}-k,\hat{Y}-l)\right|^2\cdot P_S(\hat{X}-k,\hat{Y}-l), \quad (3.39)$$

where the power spectrum of the scene irradiance is $P_S(X,Y)=\mathbb{E}\left\{\left|S(X,Y)\right|^2\right\}$.

A similar reasoning is followed by Vollmerhausen and Driggers [108]. Instead of averaging over all possible shifts, only shifts of the point-source are considered within a rectangular detector pixel. This leads to a weighting of the components in equation (3.39) with a sinc-function. In effect, the transfer function is obtained for a rectangle of constant irradiance in the scene. Unfortunately the result obtained by Vollmerhausen and Driggers would have been different if the rectangle undergoes a sub-pixel shift. This

transfer function is therefore of little practical value, unless the scene is known to consists of rectangular areas of equal irradiance which are aligned with the detector pixels. The simple formulation in equation (3.39) is therefore preferred. This is in fact equivalent to assuming a sinc-interpolation of a sampled-scene, probably a better approximation for most scenes.

In section 2.3 it was seen that detector noise in the spatial-frequency domain can often be considered to be Gaussian distributed with zero mean, irrespective of the irradiance. The spurious response should therefore show no correlation with the detector noise, so that the root-mean-square error of the detected image can be written as $\sqrt{P_N + \overline{\varepsilon^2_{\text{spurious}}}(\hat{X},\hat{Y})}$, with P_N, the power spectrum of the noise. To calculate the root-mean-square error, the error introduced by the spurious response is thus equivalent to an additional source of noise with mean zero, and variance spatial-frequency dependent $\overline{\varepsilon^2_{\text{spurious}}}(\hat{X},\hat{Y})$. The hybrid imaging error of a sampled system can thus be calculated by multiplication of the optical-transfer function with the pixel transfer function, and including the spurious error in the noise model by evaluating equation (3.39) for a given OTF and pixel transfer function using the signal model introduced in section 2.4.

It should be noted that the aliasing error introduced by the spurious response will be perceived differently from the detector noise with the same variance. The structural similarity metric [75] used in chapter 2 could provide a closer approximation of subjective judgement of image quality; however, as already discussed in the previous chapter, computationally intensive image simulations would be required which moreover can introduce a bias for the chosen set of test scenes.

3.5.3 Under-sampling in pupil-engineered systems

In under-sampled imaging systems, the OTF influences both the signal and the noise. Under-sampling can thus be expected to have a different influence on pupil-modulated optics. E.g. no aliasing effects will be visible for a sufficiently aberrated pupil because the MTF will only have significant values at low spatial frequencies. On the other hand, the spurious response will be more important when the pupil modulation favours high spatial frequencies. This could occur for instance with an annular aperture.

Hybrid imaging systems that maintain a high resolution must employ pupil modulations that yield significant MTF values across the spatial-frequency spectrum. However, pupil modulation necessarily reduces the MTF, hence less difference can be expected between the MTF at mid and high spatial frequencies. Relative to the diffraction-limited system, the signal at medium spatial frequencies decreases more than that of the spurious components. The signal will thus decrease compared to the standard deviation of the spurious error calculated via equation (3.39). Under-sampling can therefore be expected to reduce the signal-to-noise ratio of pupil-modulated systems faster than that

of a diffraction-limited system. However, this difference is minimal considering that also the power spectrum of the scene typically decreases with spatial-frequency. The use of pupil-phase modulation should therefore not be hampered by under-sampling.

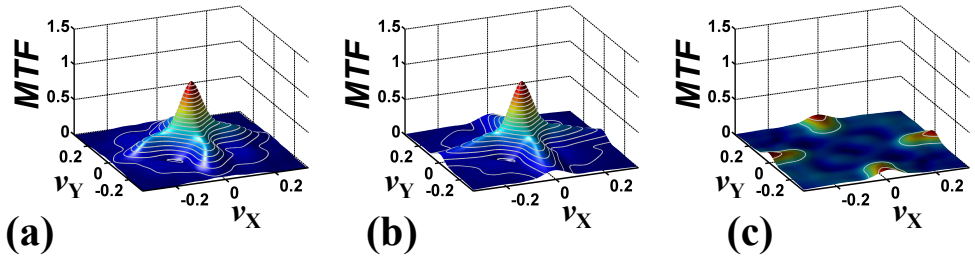


Figure 3.14: *(a) The Fourier transform of the cubic PSF shown for spatial frequencies up to the Nyquist frequency of the detector, (b) the direct response, and (c) the spurious response. The cartesian axes of the cubic phase-modulation are parallel to the detector axes.*

Figure 3.14a shows the Fourier transform of a sampled image of the point-spread function for a cubic phase-modulated imaging system, and figures 3.14b and c show respectively the direct and spurious response. The MTF can be seen to be significant only along the horizontal and vertical spatial-frequency axes, high spatial-frequency components are suppressed unless the frequency is near one of the high ridges in figure 3.14b. This asymmetry means that the orientation of the phase-modulation with respect to the sensor array has an influence on the signal-to-noise ratio, and could be exploited to reduce the impact of aliasing. A rectangular detector array permits the sampling of higher spatial frequencies in diagonal directions, it would thus be sensible to rotate the phase-modulation by $\theta = 45^\circ$ with respect to the sensor. Figure 3.15a shows the transfer function for such a rotated phase-modulation, together with its separation in a direct and spurious response shown in figures 3.15b and c respectively. It can be seen from figure 3.15b and c that the spurious response lines up with the direct response at $\theta = 45^\circ$. As for the case of $\theta = 0^\circ$ shown in figure 3.14, the spurious response perturbs the useful spatial frequencies in the direct response. As shown in figures 3.15(d-f), an intermediate rotation of $\theta = 30^\circ$ could avoid interference with the high values in the direct response; however, lower spatial frequencies in the direct response are rendered useless due to the presence of considerably larger spurious components.

To investigate this further, the hybrid imaging error is plotted in figure 3.16 as a function of the orientation θ of the cubic phase-modulation with respect to the sensor grid. The signal model of the previous chapter is used, and the spurious variance is introduced in the noise model. Figure 3.16a and b show the imaging error for a respectively three, and five times under-sampled system. The variation with θ is modest for the three times under-sampled system, nevertheless it can be seen that the hybrid imaging error is minimised at $\theta = 45^\circ$. Although for a five times under-sampled system the variations become more prominent, the optimum orientation remains diagonal.

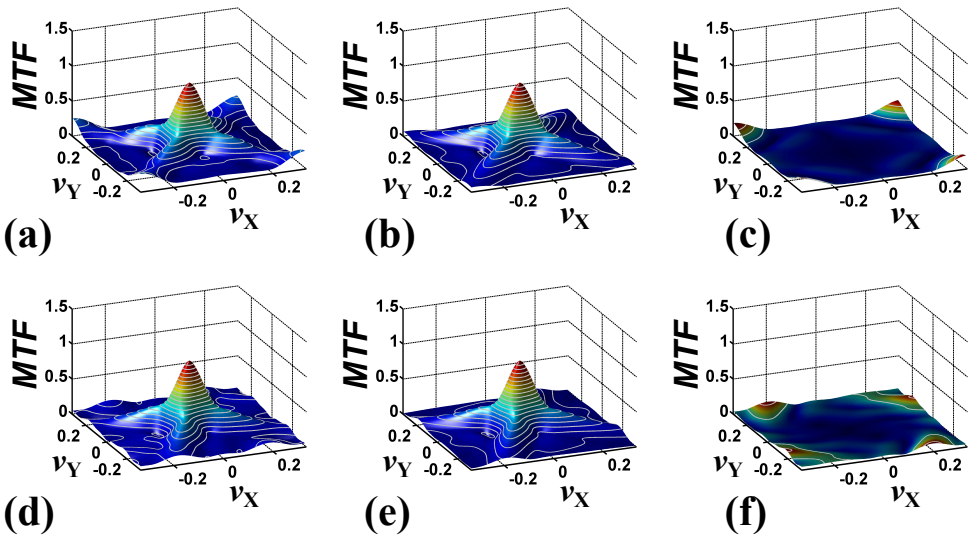


Figure 3.15: *The Fourier transform of the PSF of a rotated cubic phase-modulation, (a) by* 45°*, and (b) by* 30°*. The direct response is shown in (b) and (d), and the spurious response in (c) and (f), for the orientation of* $\theta = 45°$ *(b,c) and* $\theta = 30°$ *(e,f) respectively.*

Even for five times under-sampling, the orientation dependence is modest. This could be explained by the relatively low weight of high spatial-frequency information in typical scenes. The orientation of the phase-modulation could bear more importance when the high spatial frequencies are important to the application, e.g. if the imaging system is to be used for target localisation.

To assess the impact of under-sampling on the aberration tolerance, a hybrid imaging system with a cubic phase-modulation is evaluated for a defocus tolerance $W_{20,\max} = 2\lambda$. The phase-modulation magnitude, α, is optimised for two identical systems apart from the pixel separation. Both systems are f/1 and would thus require a pixel size of 5 µm at $\lambda = 10$ µm to avoid under-sampling. However only the first system has pixels at 5 µm, the detector of the second system has pixels separated by 25 µm and is thus five fold under-sampled. For both systems the pixel shape is simulated as a square covering 85% of the available area, a typical value for thermal imaging systems.

The hybrid imaging error is calculated using a peak-SNR of 100, and a signal that decreases with spatial-frequency f [mm] as $0.6/f$. Although the imaging error for the under-sampled system is considerably larger due to the loss of high spatial frequencies, the phase-modulation amplitude α is found to be remarkably similar for both systems as can be seen from table 3.2.

To simulate the optimal image restoration algorithm, *a priori* knowledge of the defocus has been assumed for the simulation results of table 3.2. When the optimisation is repeated using a single Wiener filter corresponding to the OTF at $W_{20} = 0$, the optimal values of table 3.3 are found. An image-shift is introduced because the cubic phase modulation results in a parabolic caustic through-focus [18, 110–113]. Such a shift does

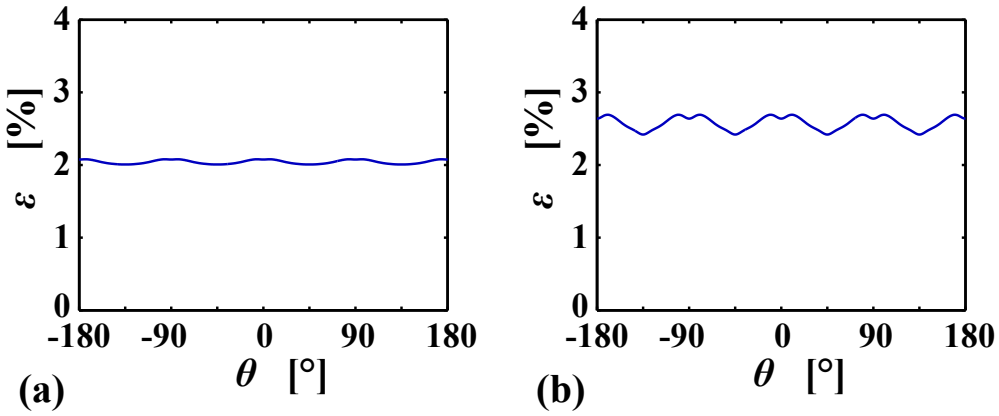


Figure 3.16: *The imaging error of a hybrid imaging system with a cubic phase-modulation under-sampled by a factor of (a) three, and (b) five; shown as a function of the orientation θ of the cubic phase-modulation in the pupil plane.*

pixel separation	defocus	ε	α
5 µm	known	1.08%	1.236λ
25 µm	known	2.95%	1.239λ

Table 3.2: *The minimum expected hybrid imaging error for a defocus tolerance $W_{20,\max} = 2\lambda$, and optimal modulation magnitude, α, for an under-sampled imaging system incorporating a cubic phase-modulation. The pixel separation of 5 µm prevents under-sampling, and the pixel separation of 25 µm is five times under-sampled. No mismatch is assumed between the restoration kernel and the actual defocus.*

not affect the perceived image quality[19], as such it is removed by co-registration. As expected, the hybrid imaging error is still somewhat larger due to the phase-mismatch between the restoration filter and the actual OTF when $W_{20} \neq 0$ [99, 47], this leads to artefacts in the restored image that look like edge-replications [47]. The somewhat larger α values found in table 3.3 counteract the phase-variance at a cost of a modest reduction of the MTF. As a result the hybrid imaging error is only marginally larger than that of the systems using the optimal image restoration algorithm. The difference between the optimal α-value is now larger than that shown in table 3.2; however, it remains sufficiently small to conclude that under-sampling has no practical influence on the defocus tolerance of the cubic phase-modulation.

Although the optimal cubic phase-modulation is almost unaffected by under-sampling, it has not been proven that this is generally the case. Other types of pupil mo-

[19]The parabolic shift with defocus does cause *unnatural* visual effects at occlusions, i.e. the background scene can be seen behind edges with a specific orientation.

pixel separation	defocus	ε	α
5 µm	**un**known	1.12%	1.564λ
25 µm	**un**known	3.28%	1.588λ

Table 3.3: *The minimum expected hybrid imaging error for a defocus tolerance $W_{20,\max} = 2\lambda$, and optimal modulation magnitude, α, for an under-sampled imaging system incorporating a cubic phase-modulation. Image restoration is done irrespective of the actual defocus using the Wiener filter calculated for the OTF at $W_{20} = 0$.*

dulation could respond differently to under-sampling, and even for the cubic phase-modulation a modest increase in the α-value is found. A rigourous evaluation and optimisation of a hybrid imaging system must therefore take the effects of sampling into account. This is demonstrated with several examples for realistic optical designs in chapter 4.

3.6 Conclusion

This chapter discussed the method for PSF calculation implemented for the simulations presented in this work. Although the discussion focuses on the implementation with *Zemax*, the principles are generally applicable. An efficient method is introduced for the calculation of the monochromatic scalar field PSF of a general optical model. Although the scalar field-approximation is often sufficient, its extension to polarised illumination is straightforward. It is shown that pupil-engineered optics can have different requirements for the simulation of pan-chromatic illumination. For the case of the cubic phase-modulation it is shown that almost forty monochromatic PSF calculations would be required to guarantee an accurate simulation, where four wavelengths would suffice to simulate a diffraction-limited system. However, considerably fewer wavelengths can provide an accurate approximation of the MTF at the often more important low spatial frequencies.

The effects of sampling, and their impact on the imaging error are discussed in detail. It is shown how the imaging error can be estimated readily for an under-sampled system, based on the direct transfer function and a signal model as that introduced in chapter 2. Several implications specific to pupil-engineered imaging systems are considered. The orientation of a cubic phase-modulation with respect to the detector array is found to have an influence on the imaging error in an otherwise axially-symmetric lens system. And although the expected imaging error is larger for an under-sampled defocus-tolerant imaging system, the optimal cubic phase-modulation magnitude α is found to be largely unaffected. Nevertheless, accurate simulations are only possible by

considering the effects of sampling. The implementation of the simulations presented in chapter 4 therefore account for aliasing and the pixel transfer function.

Chapter 4 – Applications

4.1 Introduction

In this chapter, the method described in chapter 2 is applied to the design of two realistic imaging systems: a thermal imaging system with four germanium lens elements, discussed in section 4.2; and a compact, wide-field reflective system with two mirrors, discussed in section 4.3. Rigourous simulation of both designs is achieved by integration of the optical-design and ray-tracing software *Zemax* with the general purpose computing software *Matlab* to account for detector effects and image restoration as discussed in chapter 3. It is shown that although the computational time of a hybrid imaging fidelity metric is larger than that of traditional optical-design metrics, the optimisation of the pupil modulation for a particular application is already practical and can lead to novel optical-digital designs.

4.2 Comparison of common phase-modulations for extending the depth-of-field

In this section, various pupil-modulation functions are optimised and evaluated for their ability to extend the depth-of-field of a hybrid imaging system. In the first instance the phase-modulation is introduced in the pupil of a hypothetical Nyquist-sampled paraxial imaging system, and in the second instance the surface-sag of the first lens element is modulated in an uncooled long-wave infrared imaging system exhibiting significant aliasing. In contrast to the analysis in chapter 2 where an ideal image restoration process has been assumed, here both systems employ a simple Wiener filter that minimises the mean-square imaging error when the object is in focus. Nevertheless, the systems are also tested for out-of-focus planes where there is a mismatch between the PSF and the restoration kernel.

As became clear in chapter 2, the appropriate choice of phase-modulation, both in type and parameterisation, is crucial to the performance of hybrid optical-digital imaging system. The large variety of phase modulations proposed in the literature, often analysed using ad-hoc metrics – further complicates the optical-digital design process. In this section, six of the most common antisymmetric and rotationally-symmetric phase-modulation types are compared for their ability to extend the depth-of-field of hybrid optical-digital imaging systems. To permit a valid comparison, all phase-modulation types are optimised for the metric used in the comparison. As will be shown in this section, three out of four radially-symmetric phase modulations converge to a similar pupil-phase profile, and hence expected imaging error. The cubic phase modulation yields the lowest expected imaging error; however, now that a simple Wiener filter is

used for the deconvolution, the secondary generalised cubic optimum yields a larger imaging error than the radially-symmetric modulations.

In the following subsection an overview of the phase-modulation functions is presented as well as a generalisation of some of the types, leading to simplified equations of independent parameters that facilitate the optimisation and comparison. In section 4.2.2, the hybrid imaging system and the conditions of the optimisation process are detailed. While in section 4.2.3 the outcome of the optimisations are presented and compared, showing the convergence of the radially-symmetric phase-modulations and the relative comparison of the six types of phase modulation.

4.2.1 Review of the most relevant pupil-phase modulations

Most of the reported phase-modulations can be described as, or are variations of either the antisymmetric cubic [18] with its generalisations [64, 19, 114], or the radially-symmetric log-asphere [50, 54], the axicon [51], the axilens [115], the log-axicon [53], and the quartic phase mask [57].

The first antisymmetric phase-modulation proposed to increase the defocus tolerance is the rectangularly separable cubic phase-modulation [18]:

$$\theta(x,y) \quad = \quad \alpha\left(x^3+y^3\right), \tag{4.1}$$

where x and y are normalised pupil coordinates, and α is a free parameter regulating the defocus tolerance as seen in chapter 2. The cubic phase mask may be considered as a subset of a more general family of phase-modulations described by the extended third-order polynomial:

$$\theta_{\mathrm{c}}(x,y) \quad = \quad \alpha\left(x^3+y^3\right)+\beta\left(x^2y+xy^2\right), \tag{4.2}$$

which, as seen in chapter 2, are also expected to yield good defocus tolerance when $\beta \equiv -3\alpha$, the form corresponding to trefoil. Additionally, this generalised-cubic has been shown to provide better tolerance to astigmatism [94]. The local optima with $\beta \approx 0$ and $\beta \equiv -3\alpha$ are evaluated for both hybrid imaging systems considered in this section, and are referred to as the *general cubic* and *trefoil* phase-modulation respectively.

The classical *axicon* proposed by McLeod focusses a collimated beam onto a line segment between the axicon and a point at a distance z_2 on the optical axis [51]. The phase function is given in units of wavelength by:

$$\theta(\rho) = -\frac{1}{\lambda}\left(1+\frac{z_2^2}{R^2}\right)^{-1/2} R\rho, \tag{4.3}$$

where ρ is the normalised radial coordinate, and R the aperture radius. Within the context of geometrical optics, it can be seen that the intensity is concentrated along the axis and increases linearly between the tip of the axicon and the point at z_2. In conjunction with a focussing lens, this extended depth-of-focus could also be used to extend the depth-of-field of an imaging system. To permit fine-tuning the focal distance in object space, all constants are combined into a single parameter, p_{ax1}, and introduce a quadratic term in the axicon equation:

$$\theta_{\mathrm{ax}}(\rho) = p_{\mathrm{ax1}}\rho + p_{\mathrm{ax2}}\rho^2, \tag{4.4}$$

where the independent parameters p_{ax1} and p_{ax2} are to-be-optimised for the same metric used to analyse the other phase-modulations.

A more practical phase function to extend the depth-of-field could be the axilens which yields a uniform axial intensity distribution between two points at focal distances z_1 and z_2 [115]. Sochacki et al. showed that in the paraxial approximation such an intensity distribution can be obtained by introducing an optical-path-difference of [53]:

$$\theta(\rho) = -\frac{1}{\lambda}\frac{R^2}{2\Delta z}\ln\left(z_1 + (z_2 - z_1)\,\rho^2\right), \tag{4.5}$$

in units of wavelength. Since the term axilens has been used to refer to different pupil functions with similar properties [115], the phase function derived by Sochacki et al. is called here the *logarithmic axicon* in reference to its logarithmic profile. It can be seen that equation (4.5) can be rewritten as a function of two free variables p_{lax1} and p_{lax2}, and generalised by adding a refocussing term $p_{\mathrm{lax3}}\rho^2$:

$$\theta_{\mathrm{lax}}(\rho) = p_{\mathrm{lax1}}\ln\left(p_{\mathrm{lax2}}\rho^2 + 1\right) + p_{\mathrm{lax3}}\rho^2, \tag{4.6}$$

where p_{lax1} to p_{lax3} are variables that can be optimised independently.

Using diffraction theory, Zalvidea et al. derived the *quartic* phase modulation that yields a high through-focus, on-axis intensity [57]. The derivation of the phase profile is based on the relationship between the Wigner distribution function associated with the pupil function and the on-axis intensity. The quartic phase modulation is described as:

$$\theta_{\mathrm{q}}(\rho) \;=\; \gamma\left(\rho^2 + \delta\right)^2, \tag{4.7}$$

with $\delta = -1/2$, and the parameter γ controls the defocus tolerance. Here, both variables are optimised independently. Zalvidea et al. showed that with equivalent magnitude of the phase transmission, the on-axis irradiance distributions of the quartic phase modulation and logarithmic axicons are practically the same, a result which is confirmed in what follows.

Chi and George showed that on-axis points in object space can be projected to a single point in image space using a lens with radially varying focal length [50], that has been named the *logarithmic asphere*. At any object distance within the design range there is always a circular annulus that contributes to sharp imaging, while the remainder of the lens causes blur. Digital processing of the intermediate blurred image caused by the logarithmic asphere yields sharp images with extended-depth-of-field.

The logarithmic asphere for typical f-numbers is given in units of wavelength by [50]:

$$\theta(\rho) \;=\; -a_\beta \frac{t^2+\rho^2}{2} \ln\left[A_\beta\left(t^2+\rho^2\right)-1\right], \tag{4.8}$$

where the constants a_β, and A_β are given by the equations:

$$a_\beta \;=\; \frac{1}{\lambda_0} \frac{1/s_1 - 1\Big/\sqrt{R^2+s_2^2}}{\ln\left[\left(R^2+t^2\right)/t^2\right]} \tag{4.9}$$

$$A_\beta \;=\; \frac{1}{t^2} \exp\left[-\frac{\sqrt{R^2+s_2^2}}{\sqrt{R^2+s_2^2}-s_1} \ln\left(\frac{R^2+t^2}{t^2}\right)\right], \tag{4.10}$$

where the constants s_1, s_2, t, λ_0, and R are problem specific. Although these constants are well-defined to optimise the criterion set by Chi and George, we replace the constants by five independent variables $p_{\text{las}1\ldots5}$, to permit optimisation for the same metric used to evaluate and optimise the other phase-modulation types:

$$\theta_{\text{las}}(\rho) \;=\; \left(p_{\text{las}1}+p_{\text{las}2}\rho^2\right) \ln\left(p_{\text{las}3}\rho^2+p_{\text{las}4}\right)+p_{\text{las}5}\rho^2, \tag{4.11}$$

where the quadratic term has been added at the end to permit refocussing and optimally cover the defocus range of interest.

4.2.2 Optimisation and evaluation of phase-modulations

The various phase-modulation types are evaluated for their ability to extend the depth-of-field of two systems; an ideal system combining paraxial optics with a Nyquist-sampled detector, and the uncooled f/1 thermal infrared imaging system shown in figure 4.1, operating at wavelengths between 8 µm and 12 µm, under-sampled with a pixel pitch of 25×25 µm, and a fill-factor of 85%. Focused at the hyperfocal distance of 225 m, this imaging system is able to keep objects from 112.5 m to infinity in focus. However, a significantly larger depth-of-field is required, so that sharp focus can be obtained for object distances from 17.5 m to infinity. After refocusing the lens at 35 m, this corresponds to $W_{20} \approx -2\lambda_0$ to $W_{20} \approx +2\lambda_0$, at $\lambda_0 = 10$ µm. Pupil-phase modulation

is achieved by modulating the front surface of the first lens-element following one of the equations given in the previous section. This can be implemented in *Zemax* with either an extended polynomial or an extended odd-asphere surface type. A least-squares fit of a 25^{th}-order odd-asphere is used to model the phase functions that contain a logarithm: $\theta_{\mathrm{lax}}(\rho)$ and $\theta_{\mathrm{las}}(\rho)$.

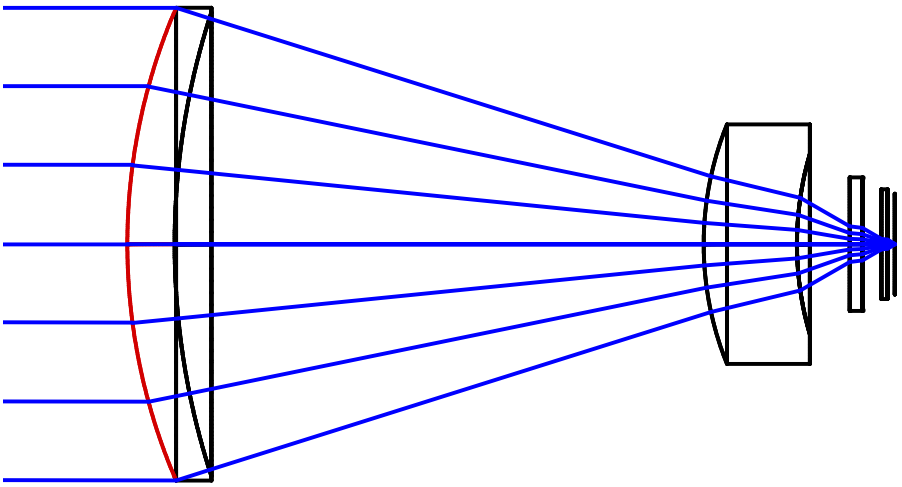


Figure 4.1: *A model of a germanium lens incorporating a phase-modulation in the front surface, marked in red, at the aperture stop. The edges of the lens elements and the image plane are shown in black. Seven rays emanating from an on-axis point source are shown; however, a square grid of* 128×128 *rays is used to estimate the pupil function.*

The parameters of the six phase-modulation types introduced previously are optimised to reduce $\bar{\varepsilon}$, the root-mean-square difference between the geometric projection of the scene and the restored image of the hybrid imaging system in which the phase-modulation is incorporated. This difference is dependent on the scene that is imaged; however, as shown in chapter 2, an appropriate estimate can be obtained with the help of the statistical model. The expected imaging error is calculated using a conservative peak signal-to-noise ratio of 100, which is also used to calculate the Wiener filter that minimises the mean-square-error between the in-focus image and the geometrical image. Consequently the optical system is characterised with its PSF at the distances 17.5 m, 20 m, 23 m, 35 m, 70 m, 140 m, and $+\infty$, or equivalently for the paraxial evaluation at $W_{20} = -2\lambda$, -1.5λ, -1λ, 0, 1λ, 1.5λ, and 2λ. This allows the estimation of the hybrid imaging error for the extended-depth-of-field. Note however that the chosen depth-sampling gives relatively more importance to the images with higher amounts of defocus to reflect the requirement of acceptable performance at distances close to the imaging system.

As is typical for most uncooled thermal imaging systems, the system under consi-

deration is sampled well below the Nyquist frequency. To account for the imaging error due to aliasing, the optical-transfer function is first calculated without under-sampling, and weighted using the statistical model of the scene [84, 85]. As has been shown in section 3.5, this permits the calculation of the spurious response as a function of spatial-frequency when the system is under-sampled. The finite area of sensitivity of each pixel is modelled with a square area of sensitivity covering 85% of the total pixel area; and the polychromatic Huygens OTF was calculated in closed-loop with *Zemax* and *Matlab*, as discussed in the previous chapter.

Although the cubic PSF is approximately invariant as a function of defocus, it is well known that it follows a parabolic transverse displacement with defocus [112, 116]. This causes an image shift when using a linear deconvolution algorithm for the image restoration. For most applications an image shift smaller than a few pixels is irrelevant, hence this shift is ignored in this section by applying co-registration of the images before the evaluation. Applications that do require absolute localisation require a larger cubic phase-modulation to minimise the image shift [18]; however, a larger cubic modulation is associated with a larger expected imaging error as has been shown in chapter 2.

4.2.3 Optimisation results and discussion

Each phase-modulation type is optimised for the conditions described in section 4.2.2, using a combination of global optimisation with the differential evolution algorithm [117], and local optimisation with the Nelder-Mead simplex algorithm [118]. The results of this optimisation are listed in table 4.1 for the Nyquist-sampled paraxial imaging system.

From the last column it can be seen that the cubic phase-modulation with a modest cross-term $\beta = 0.20\lambda$, yields the lowest imaging error, $\bar{\varepsilon} \approx 1.12\%$. The α-value of 1.54λ is somewhat larger than expected ($\alpha \approx 1.3\lambda$ for $W_{20,\max} = 2\lambda$ as can be read from figure 2.10a); however, this could be explained by the relative importance that has been given to large W_{20} by the selection of the distance-samples. Note that the pure cubic with $\beta \equiv 0$, not shown here, has a marginally higher imaging error; however, its expected imaging error remains lower than that of the other modulation types considered here.

A larger discrepancy is found for the trefoil modulation, its relatively large modulation depth of $\alpha = 1.16\lambda$ as compared to $\alpha \approx \lambda/2$ found earlier, cannot be explained by the non-uniform importance of the defocus values. There are two additional differences between this system and the ideal system studied in chapter 2: the peak signal-to-noise is 100 instead of 280, and a single digital restoration kernel is used independently of the actual defocus. By repeating the simulations with a peak signal-to-noise of 280 it was verified that the higher noise level does not lead to a larger α-value. On the other

phase-modulation type	parameters	$\bar{\varepsilon}$ [%]
cubic, global optimum	$\alpha = 1.54\lambda$, $\beta = 0.20\lambda$	1.120
trefoil	$\alpha = 1.16\lambda$, $\beta \equiv -3\alpha$	1.221
axicon	$p_{\text{ax1}} = -3.46\lambda$, $p_{\text{ax2}} = 3.41\lambda$	1.188
logarithmic axicon	$p_{\text{lax1}} = 63.17\lambda$, $p_{\text{lax2}} = 0.35$, $p_{\text{lax3}} = -19.57\lambda$	1.139
quartic	$\gamma = 2.89\lambda$, $\delta = -0.42$	1.139
logarithmic asphere	$p_{\text{las1}} = 59.95\lambda$, $p_{\text{las2}} = 20.10\lambda$, $p_{\text{las3}} = 0.23$, $p_{\text{las4}} = 0.73$, $p_{\text{las5}} = -14.95\lambda$	1.139

Table 4.1: *Optimal parameters and minimal expected imaging error for various types of phase-modulation introduced at the pupil of a paraxial imaging system with a Nyquist-sampling detector.*

hand, if the restoration kernel is calculated for an *a priori* known defocus, the much lower optimal value, $\alpha \approx \lambda/2$, expected from chapter 2 is found. It can thus be concluded that the mismatch between the digital restoration kernel and the out-of-focus PSF plays a significant role for the trefoil phase modulation. In contrast to the pure cubic where such a mismatch results mainly in an image shift, for the trefoil modulation this causes artifacts that degrade the image. The kernel mismatch, and hence the artifacts can be reduced by increasing the modulation depth, α; however this comes at the cost of a higher $\bar{\varepsilon}$. It can be seen from table 4.1 that the radially-symmetric phase modulations yield a lower $\bar{\varepsilon}$ than the trefoil modulation, in contrast to what would be expected when using an ideal image restoration algorithm as seen in chapter 2, figure 2.10c.

Although the descriptions of the radially-symmetric phase modulations are clearly distinct, it can be noted from table 4.1 that all but the axicon converge to approximately the same expected imaging error. This should be of no surprise since the associated profiles practically coincide as can be seen from figure 4.2a. Note that the optimal phase functions for the realistic under-sampled thermal imaging system shown in figure 4.2b are discussed further on.

As can be seen from figure 4.2a, apart from the classic axicon, the radially-symmetric

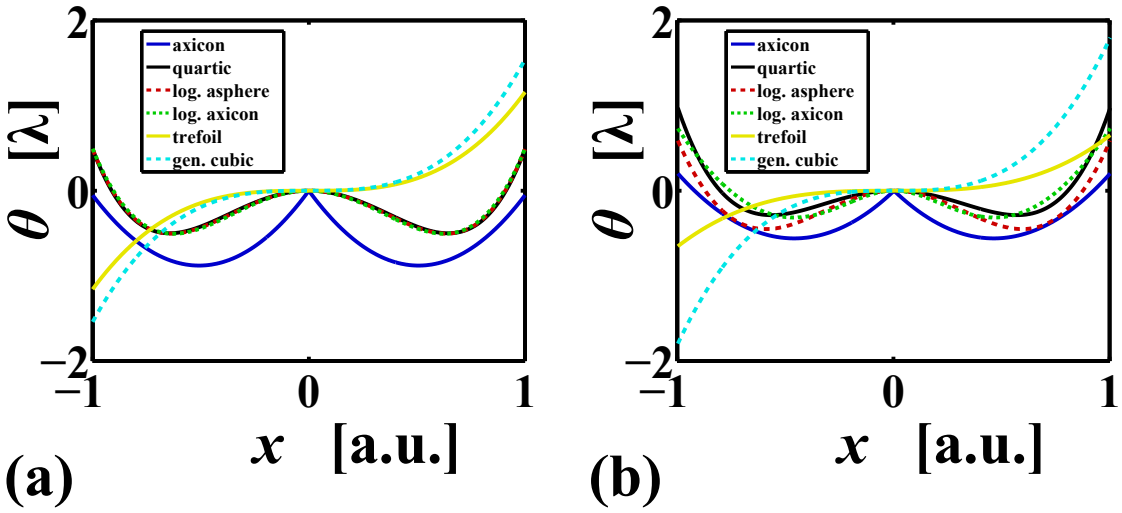


Figure 4.2: *The optimal phase-modulation profiles shown as a function of the normalised pupil coordinate x, for (a), a Nyquist-sampled paraxial system, and (b), a five-times under-sampled thermal imaging system consisting of four Germanium lens elements. The surface-sag modulation of the thermal lens is converted to equivalent optical-path difference at the central wavelength $\lambda = 10\,\mu m$. To facilitate comparison, the piston is adjusted so that $\theta(0) \equiv 0$, and the values of the logarithmic axicon profile are inverted in (a), while the values of the axicon and the logarithmic asphere are inverted in (b). This is justified since negative and positive defocus are given equal weight in the comparison, hence complex conjugation of the pupil function has no influence on the figure of merit.*

phase-modulations have practically identical optimal profiles and hence expected imaging errors, $\bar{\varepsilon} \approx 1.139\%$. One could therefore consider optimising a radial phase-modulation defined by an, often more practical, low-order radial polynomial function. The optimisation of a radially-symmetric mask where the radial phase function is defined by a tenth-order polynomial led to a modest further reduction of the imaging error to $\bar{\varepsilon} \approx 1.1320\%$.

Figure 4.3a show the MTFs of a traditional imaging system, before image restoration (dashed lines), and after image processing (solid lines). Some contrast enhancement can be noticed at low spatial frequencies for the in-focus MTF (black line); however, the contrast is lost at most spatial frequencies when the system is out of focus (coloured lines).

The same plots are shown in figure 4.4 for the six optimised phase-modulation functions. The reduction in resolution can be clearly seen when comparing the in-focus MTFs (black lines) in figure 4.4 to those of the non-modulated system shown in figure 4.3a. However, the out-of-focus MTF values are considerably higher for the pupil-modulated systems as can be noticed from figure 4.4, in particular the MTF of the generalised-cubic remains high for low spatial frequencies as can be seen in figure 4.4a.

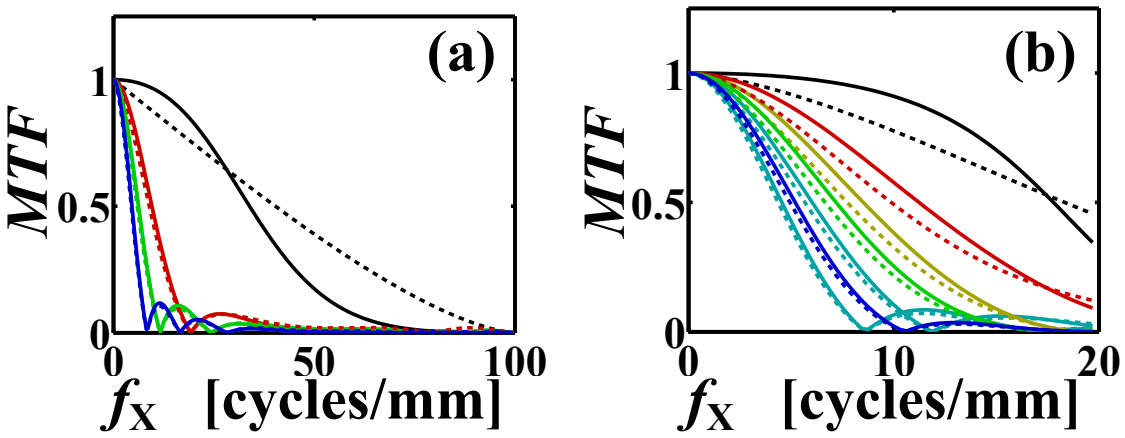


Figure 4.3: *The MTFs of a traditional imaging system, before image restoration (dashed lines), and after image processing (solid lines), shown for (a) a Nyquist-sampled paraxial imaging system, and (b) an under-sampled thermal imaging system. The colours indicate the amount of defocus (a) or object distance (b):* $W_{20} = -2\lambda$ *or* $17.5\,m$ *(light blue),* -1.5λ *or* 20 *(purple),* -1λ $23\,m$ *(dark yellow),* 0 *or* $35\,m$ *(black),* 1λ *or* $70\,m$ *(red),* 1.5λ *or* $140\,m$ *(green), and* 2λ *or* $+\infty$ *(blue).*

Although compared to the generalised-cubic, the MTF values of the other modulations are lower and vary more with defocus, they remain larger than those of the MTF of the non-modulated system depicted in figure 4.3a. Note that due to the kernel mismatch the restored modulation function is larger than unity for certain phase modulations and defocus. This could cause artifacts such as irradiance oscillations near edges in the scene; fortunately, the visual impact of this error appears to be modest here.

One-dimensional modulation transfer functions do not fully describe rotationally-asymmetric systems such as those incorporating antisymmetric phase-modulations, or as seen in section 3.5.3, under-sampled imaging systems. Image simulations permit a more complete, albeit more subjective, evaluation of the influence of pupil-modulation and image restoration on image quality.

The false-colour images in figures 4.5 and 4.6 show the difference between the geometric projection of a spoke target and the restored image simulated for the various pupil-modulations and amounts of defocus. Dark values indicate a large difference, and colours indicate the magnitude and sign of the difference. For this imaging system, the detector is assumed to point-sample the optical image at the Nyquist spatial-frequency, and 1% Gaussian noise is added before applying the regularised Wiener filter. From the central row with in-focus images it can be seen that the traditional imaging system has the lowest imaging error in-focus; however, dark values in the same column indicate a large imaging error when the traditional imaging system is out-of-focus. As can be seen from the second column, the imaging error for large defocus, and the variation with defocus is significantly reduced when using the cubic phase modulation. The variation with

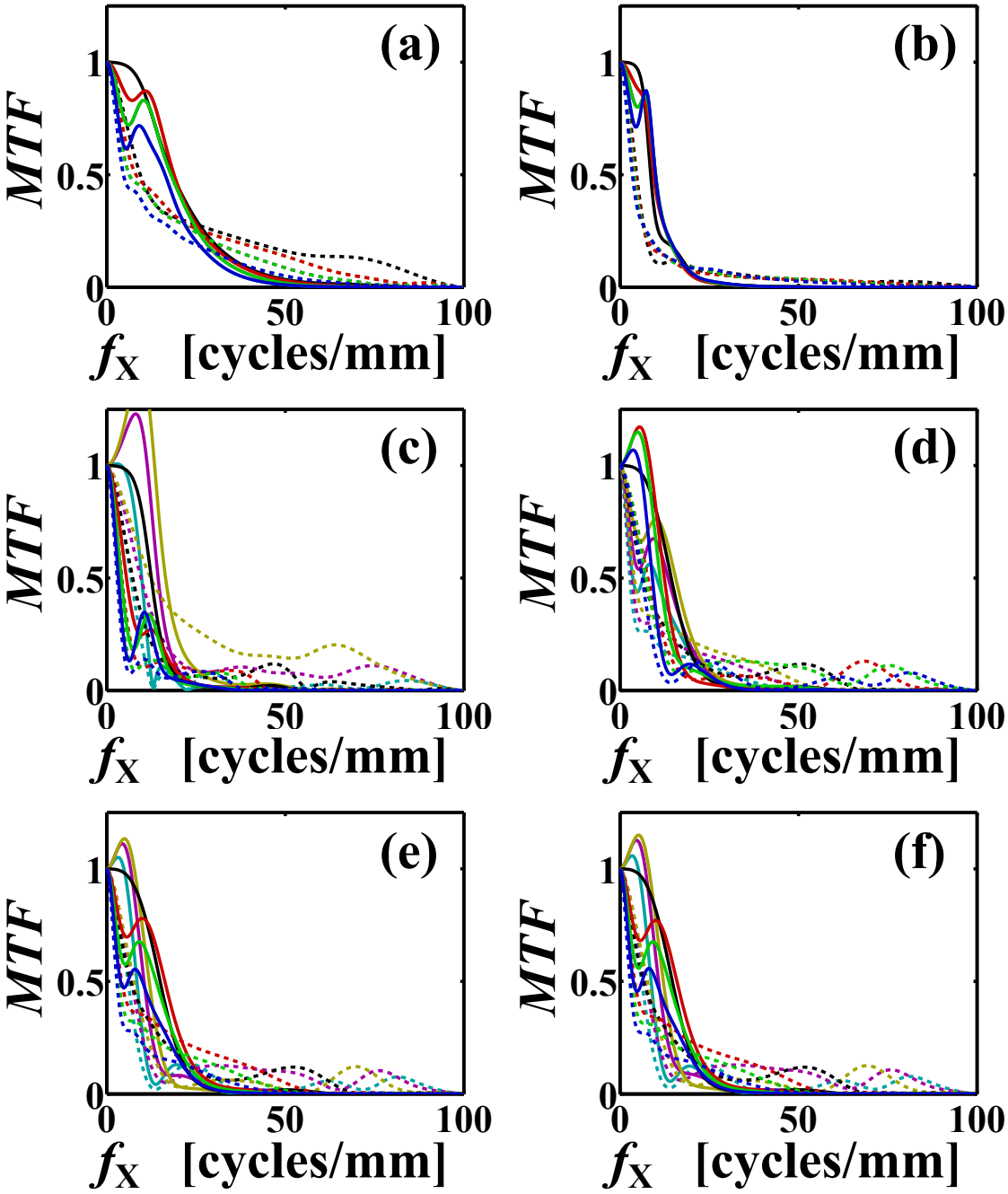


Figure 4.4: *The modulation transfer functions of a Nyquist-sampled paraxial imaging system shown for various types of pupil-phase modulation after optimisation: (a) generalised-cubic, (b) trefoil, (c) axicon, (d) logarithmic axicon, (e) quartic, and (f) the logarithmic asphere. The MTF is shown for various amounts of defocus, before deconvolution (dashed lines), and after deconvolution (solid lines). The colours indicate the amount of defocus: $W_{20} = -2\lambda$ (light blue), -1.5λ (purple), -1λ (dark yellow), 0 (black), 1λ (red), 1.5λ (green), and 2λ (blue).*

defocus is also reduced for the trefoil modulation; however, overall the imaging error is considerably larger than that associated with the cubic modulation. Three dominant directions can also be noticed in the imaging error of the trefoil modulation, leading to

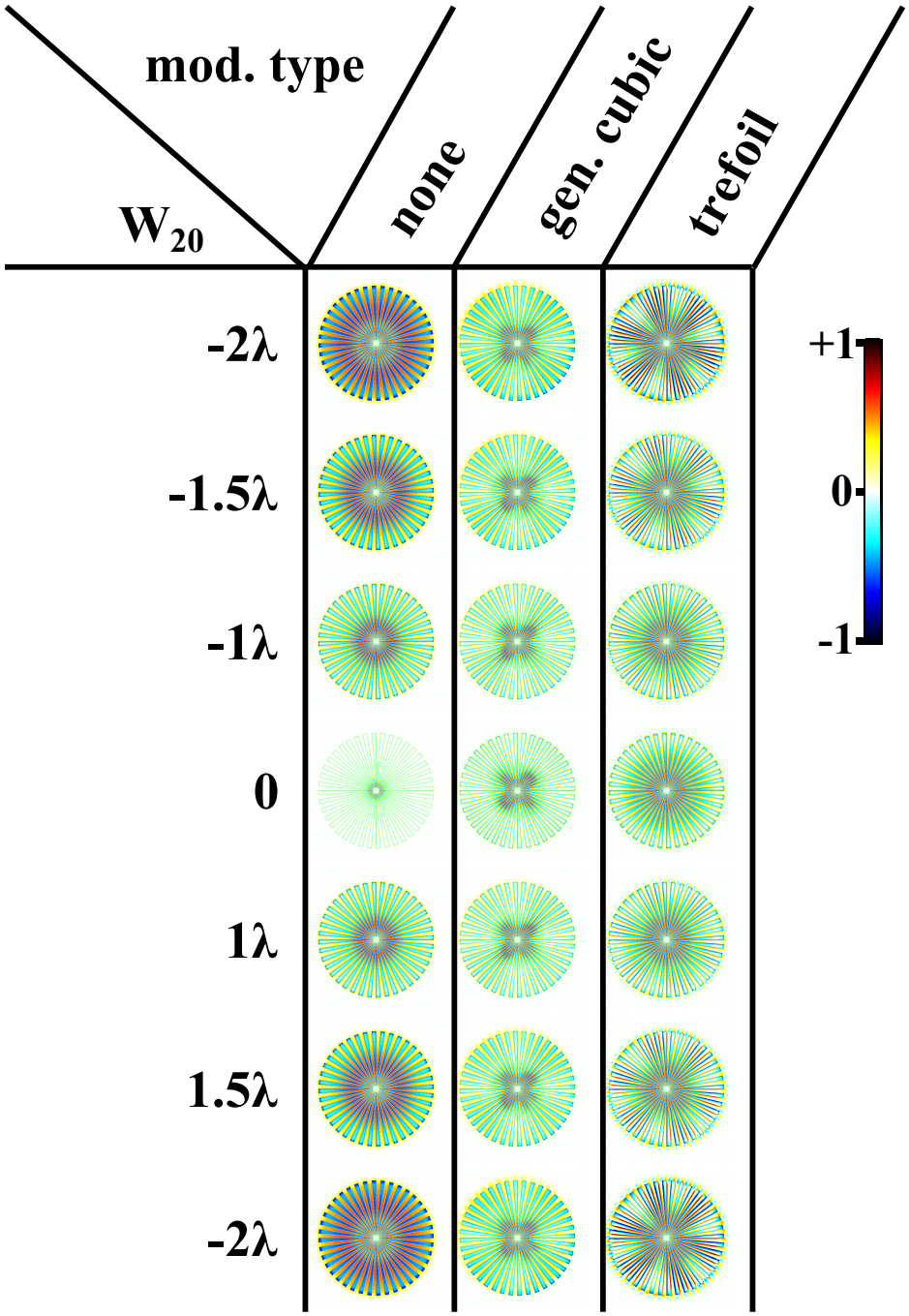


Figure 4.5: *The false-colour difference between the restored* 512×512*-pixel image of a spoke target and its geometrical image, simulated for a Nyquist-sampled paraxial imaging system. The spoke target is shown for various defocus (from top to bottom:* $W_{20} = -2\lambda$, -1.5λ, -1λ, 0, 1λ, 1.5λ, 2λ*), and for various phase-modulation types (left to right). Gaussian noise with a standard deviation of* 1% *of the dynamic range is added to the simulated images before restoration with a standard Wiener deconvolution corresponding to the in-focus transfer function. Notice that the green hues are indicative of the colour mixing due to the proximity of positive (yellow) and negative (blue) errors.*

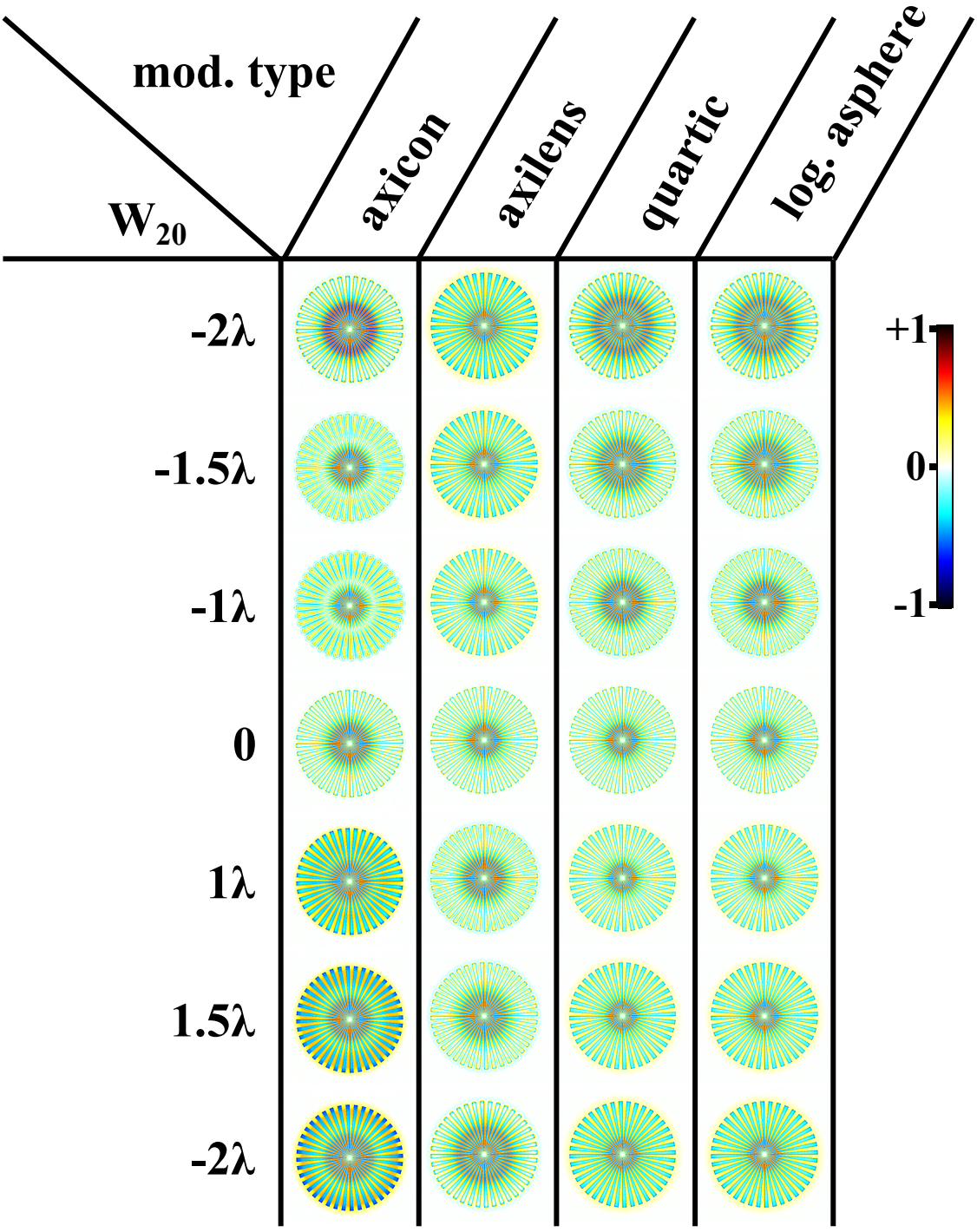


Figure 4.6: *The false-colour difference between the restored* 512×512*-pixel image of a spoke target and its geometrical image, simulated for a Nyquist-sampled paraxial imaging system. The spoke target is shown for various defocus (from top to bottom:* $W_{20} = -2\lambda,\ -1.5\lambda,\ -1\lambda,\ 0,\ 1\lambda,\ 1.5\lambda,\ 2\lambda$*), and for various radial-symmetric phase-modulation types (left to right). Gaussian noise with a standard deviation of* 1% *of the dynamic range is added to the simulated images before restoration with a standard Wiener deconvolution corresponding to the in-focus transfer function. Notice that the green hues are indicative of the colour mixing due to the proximity of positive (yellow) and negative (blue) errors.*

triangular artifacts due to the mismatch between the restoration kernel and the out-of-focus PSF. Notice that the imaging error is considerably lower for edges in the directions orthogonal to those with large imaging error. Imaging systems incorporating the trefoil modulation can thus be expected to enable the accurate determination of the position of a point-source, independently of defocus.

Due to their asymmetry, the generalised-cubic and the trefoil modulation perform equally well for positive and negative defocus. The same is not the case for the four radially-symmetric phase-modulations used for the simulations shown in figure 4.6. Even using false-colour coding, it is difficult to judge from the spoke target which of the radially-symmetric types is optimal. It can be noticed however that the axilens, the quartic, and the logarithmic asphere have similar imaging errors. Since positive and negative defocus are given equal importance, the optimisation process has resulted in an optimum for the axilens that is practically the inverse function of that of the quartic and the logarithmic asphere. Notice also that the axicon has a large imaging error at $W_{20} = -2\lambda, 1\lambda, 1.5\lambda$, and 2λ, explaining the larger expected imaging error listed in table 4.1.

phase-modulation type	**parameters**	$\bar{\varepsilon}$ **[%]**
cubic, global optimum	$\alpha = 5.9876\,\mu\text{m}$, $\beta = 2.7170\,\mu\text{m}$	3.1268
trefoil	$\alpha = 2.1758\,\mu\text{m}$, $\beta \equiv -3\alpha$	3.6664
axicon	$p_{ax1} = 8.0880\,\mu\text{m}$, $p_{ax2} = -8.7631\,\mu\text{m}$	3.4983
logarithmic axicon	$p_{lax1} = -3.3988\,\mu\text{m}$, $p_{lax2} = 6.8566$, $p_{lax3} = 9.4435\,\mu\text{m}$	3.3667
quartic	$\gamma = 9.1619\,\mu\text{m}$ $\delta = -0.3230$	3.2329
logarithmic asphere	$p_{las1} = 20.2024\,\mu\text{m}$, $p_{las2} = -8.1035\,\mu\text{m}$, $p_{las3} = -1.7516$, $p_{las4} = -1.8187$, $p_{las5} = -5.2752\,\mu\text{m}$	3.2384

Table 4.2: *Comparison of different types of phase-modulations introduced in a compound lens of a thermal imaging system.*

When the optimisation process is repeated for the under-sampled thermal imaging system shown in figure 4.1, the phase profiles depicted in figure 4.2b are obtained. It

can be noticed that the optimal phase functions are comparable to those found for the Nyquist-sampled paraxial system and depicted in figure 4.2a. Even the magnitude of most functions is comparable after conversion of the surface-sag of the germanium lens element with a refractive index, $n \approx 4.00$ at $\lambda = 10\,\mu\text{m}$, to equivalent optical-path-difference in units of wavelength. The only clear exception is the trefoil modulation which with $\alpha = 2.1758\,\mu\text{m}$ or 0.65λ is only half as large as the value found for the Nyquist-sampled system. The optimal alpha value does remain a fraction of a wavelength larger than that of the optimum predicted for use with the optimal image restoration. A more modest decrease is seen for the axicon modulation depth which is reduced from $|\gamma| = 3.46\lambda$ to $8.0880\,\mu\text{m}$ or 2.43λ, while the cubic modulation depth increased from $\alpha = 1.54\lambda$ to $5.9876\,\mu\text{m}$ or $\alpha = 1.80\lambda$. The maximum modulation of the logarithmic axicon, quartic, and logarithmic asphere remain approximately the same; however, their respective profiles do not coincide perfectly anymore as can be seen from figure 4.2b. As can be seen from table 4.2, the difference between expected imaging error $\bar{\varepsilon}$ for the quartic and the logarithmic asphere is however negligible. That of the logarithmic axicon is somewhat larger, which could suggest that the optimisation process has terminated in a local optimum. Indeed, it was already seen that the form of the logarithmic axicon can approach that of the quartic phase-function, hence the logarithmic axicon should be able to attain an expected imaging error as low as that of the quartic. The calculation of the merit-function involves an approximation of the point-spread function by tracing a grid of 128×128 rays through the pupil. The numerical error introduced in the calculations can create local optima in an otherwise constant merit function space. Furthermore, the convergence is slowed by the highly non-linear dependency between the phase-modulation parameters and the PSF. Due to the logarithm in the equations of the logarithmic axicon and the logarithmic asphere, small perturbations in the parameter values can cause significant changes in the optical-path difference. The quartic phase modulation has a more direct dependence on its parameters, and that is perhaps why it yielded the lowest imaging error after optimisation. This suggests that other radially-symmetric phase modulations could be replaced by the quartic, if necessary, amended with higher order terms.

Figure 4.3b shows the magnitude of the system transfer functions of the thermal imaging system without pupil-modulation for various object distances (colours), before (dashed lines), and after image processing (solid lines). The transfer function is shown up to the maximum spatial-frequency that can be sampled with a pixel pitch of $25\,\mu\text{m}$, and it incorporates the contrast reduction due to a finite fill-factor of 85%. It can be seen that sufficient signal is expected for all spatial frequencies with the exception of these near the cut-off spatial-frequency of the detector. As for the Nyquist-sampled MTFs, the first zero occurs at $f_X \approx 9$ cycles/mm for the nearest object position. This is remedied by the introduction of a phase modulation at the aperture stop as can be seen from the

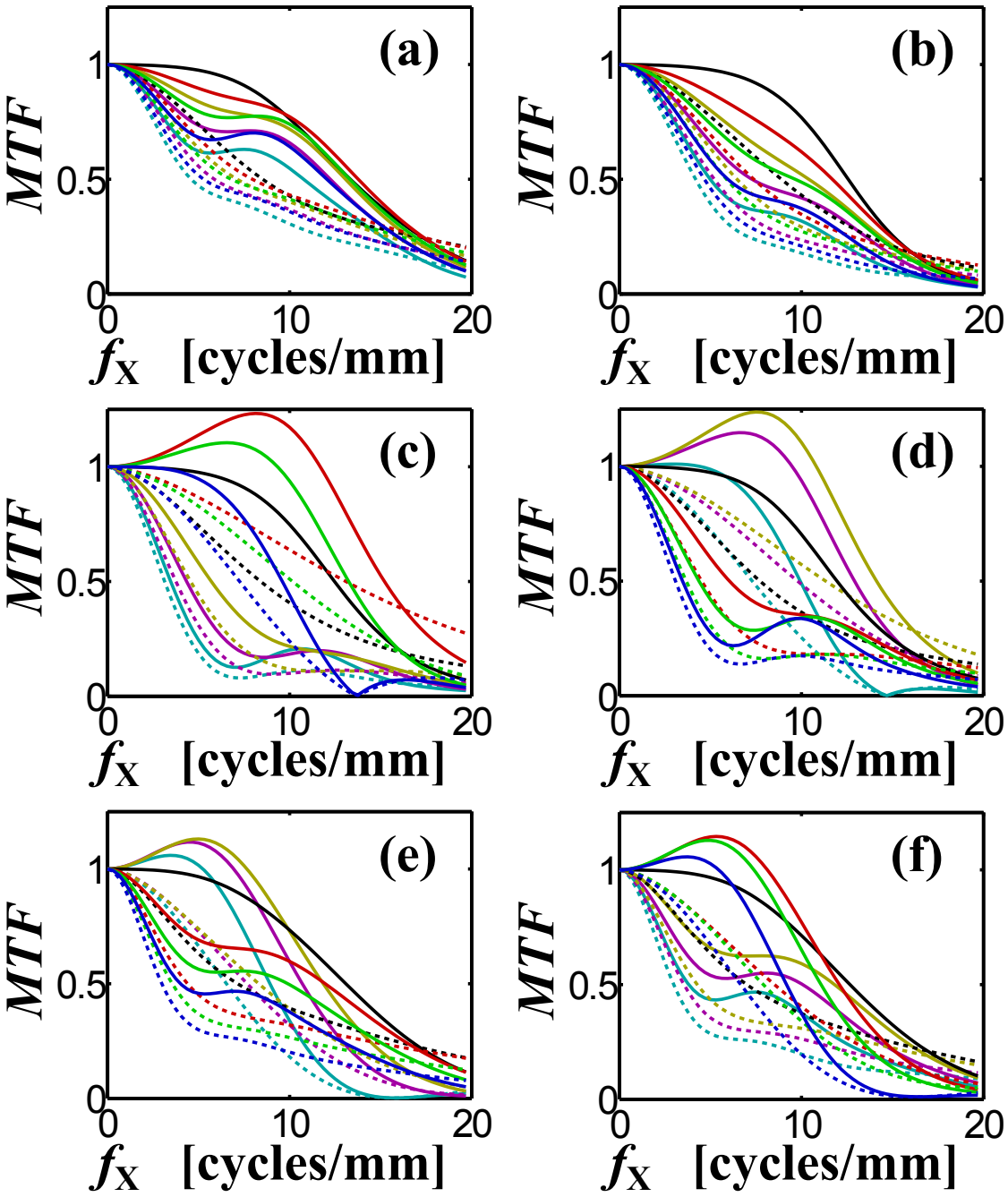


Figure 4.7: *The modulation transfer functions of a f/1 thermal imaging system with a pixel pitch of* 25 µm *shown for various types of pupil-phase modulation after optimisation: (a) generalised-cubic, (b) trefoil, (c) axicon, (d) logarithmic axicon, (e) quartic, and (f) the logarithmic asphere. The MTF is shown for various object distances, before deconvolution (dashed lines), and after deconvolution (solid lines). The line colours indicate the object distances:* 17.5 *m (light blue),* 20 *m (purple),* 23 *m (dark yellow),* 35 *m (black),* 70 *m (red),* 140 *m (green), and* ∞ *(blue).*

plots in figure 4.7. The MTF associated with the generalised-cubic and the trefoil phase modulation seen in figure 4.7a and 4.7b are free of nulls. The radially-symmetric phase modulations show more variation with defocus, and are unable to prevent nulls in the

transfer function; however, the nulls occur only for $f_X \gtrapprox 12$ cycles/mm and the extreme values of the object distance.

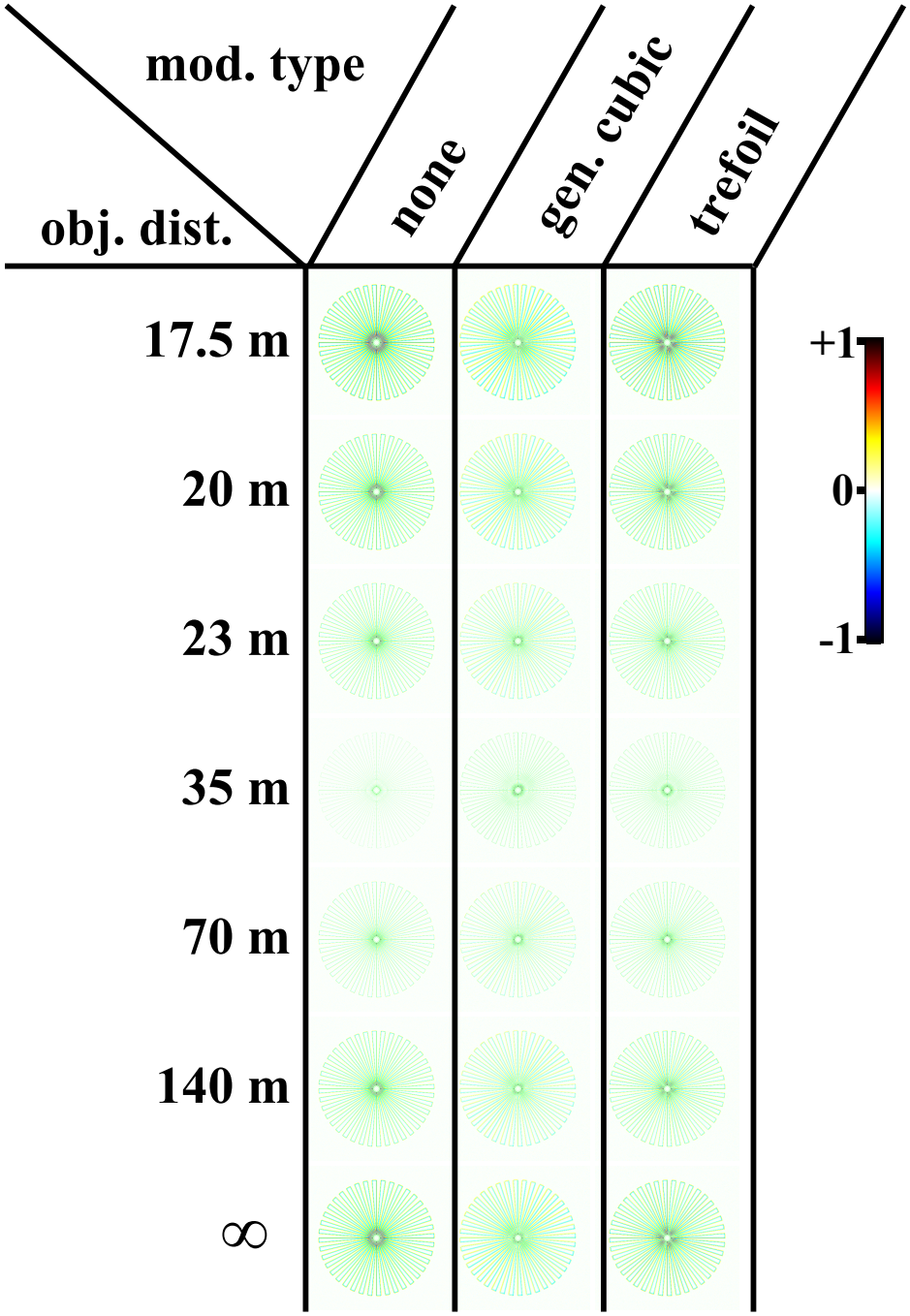


Figure 4.8: *The false-colour difference between the restored* 512 × 512*-pixel image of a spoke target and its geometrical image, simulated for an under-sampled thermal imaging system. Gaussian noise with a standard deviation of* 1% *of the dynamic range is added to the simulated images before restoration with a standard Wiener filter. Note that since the display pixel size is identical to that used in figure 4.5, the angular size of the spoke target used here is five times larger than that seen in figure 4.5. Notice that the green hues are indicative of the colour mixing due to the proximity of positive (yellow) and negative (blue) errors.*

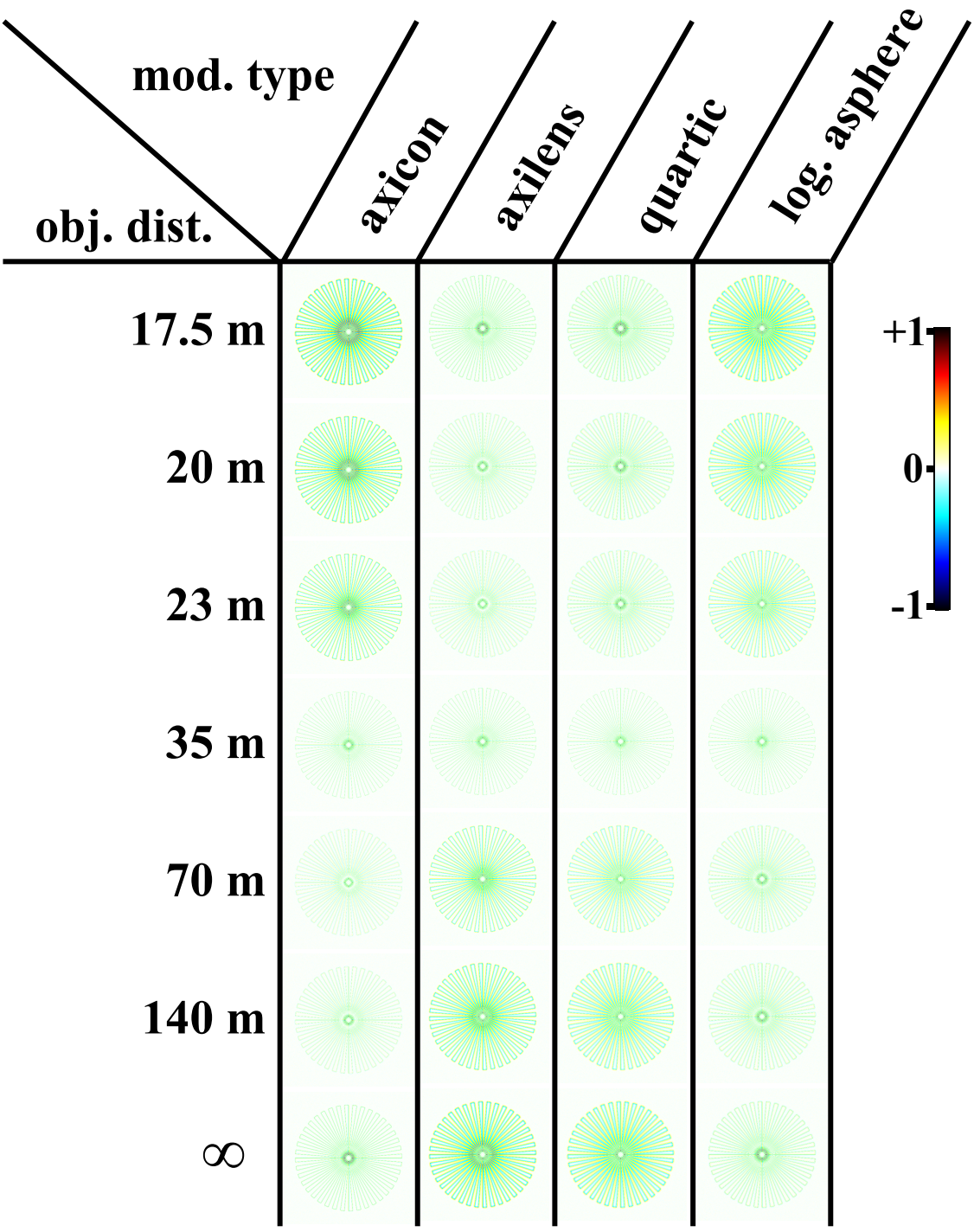


Figure 4.9: *The false-colour difference between the restored* 512 × 512*-pixel image of a spoke target and its geometrical image, simulated for an under-sampled thermal imaging system. Gaussian noise with a standard deviation of* 1% *of the dynamic range is added to the simulated images before restoration with a standard Wiener filter. Note that since the display pixel size is identical to that used in figure 4.6, the angular size of the spoke target used here is five times larger than that seen in figure 4.6. Notice that the green hues are indicative of the colour mixing due to the proximity of positive (yellow) and negative (blue) errors.*

Figures 4.8 and 4.9 show the false-colour difference between the geometrical projection of a spoke target and its image obtained with the hybrid thermal imaging system. The left-most column of figure 4.8 shows the simulation results for the thermal imaging system without pupil-phase modulation. As can be seen from the light image at the centre of the column, the expected imaging error is minimised at 35 m. The lens is designed to minimise the wavefront error at this distance so that objects at infinity are out-of-focus with $W_{20} \approx 2\lambda$. The imaging error can be seen to increase with the amount of defocus as expected. The loss of high spatial frequencies blurs the edges, and a dark ring at the centre of the top spoke target corresponds to the null in the transfer function at $f_X \approx 9$ cycles/mm. The images of the system with the generalised-cubic pupil modulation show considerably less variation in quality, only a modest blurring of the edges can be observed due to the signal suppression at the spatial frequencies near the detector cut-off.

4.2.4 Conclusion

The generalised-cubic phase modulation is found to yield the highest fidelity image, both for a Nyquist-sampled paraxial system, and for a realistic thermal imaging system with considerable aliasing. Although the cubic PSF is approximately invariant as a function of defocus, it is well known that it is displaced as a quadratic function of defocus [119]. This causes an image shift when using a linear deconvolution algorithm for the image restoration. For most applications an image shift smaller than a few pixels is irrelevant, hence the optimisation results presented in this section cancelled this shift before calculating the imaging error. Applications that do require absolute localisation would require a stronger cubic phase-modulation so that the image shift is minimised [18], this however leads to a larger imaging error as was shown in chapter 2. Alternatively, radially-symmetric phase-modulations can be used to increase the tolerance to defocus. For certain applications the complete lack of image shift might out-weight the higher imaging error that is expected for radially-symmetric phase modulations. Although the imaging error of radially-symmetric modulations shows some variation with defocus, it was seen that it can be kept to an acceptable level.

The optimal radially-symmetric phase modulation was found to be the quartic modulation; however, both the logarithmic axicon and the logarithmic asphere converged to similar phase profiles and thus comparably low imaging error. The relative simplistic formulation of the quartic is not only more practical for implementation in an optimisation process, also its cost function varies smoothly with changes in its parameters. Out of the radially-symmetric phase modulations presented here, it appears thus sufficient to evaluate only the quartic modulation, perhaps amended with higher order terms, as this was shown to further reduce the imaging error by a modest amount.

The maximum phase modulation of all types evaluated here was never found to be larger than two wavelengths to tolerate a maximum defocus of $W_{20} = 2\lambda$. Most authors however suggest the use of considerably larger phase modulations which would lead to excessive imaging errors due to a loss in resolution or noise amplification.

In this section, the thermal imaging system was only analysed for on-axis performance. It was verified that the optics are well-corrected for off-axis aberrations; and since the pupil-modulation is introduced at the aperture stop, it does not affect the translation-invariance of the PSF. Although an on-axis analysis is sufficient for this particular system, in general the image quality should be optimised for the full field-of-view as shown in the next section.

4.3 Holistic design of a wide-field all-reflective imaging system

So far our discussion has been limited to the simulation of on-axis point-sources. This is not overly restrictive as the optical system is typically designed to be approximately invariant with field direction. In some cases however it is not possible to make this assumption, e.g. conformal optics are characterised by large aberrations that can vary significantly across the field-of-view. Low-cost or simplified optics might have non-negligible variations of the point-spread function with field, in this case hybrid imaging using phase-modulations has already shown its advantages [120–122, 95].

In this section it is shown how the method proposed in this thesis can be employed to optimise a hybrid imaging system for defocus tolerance while at the same time assuring acceptable image quality across the full field-of-view. This enables the design of a compact wide-field reflective imaging system that provides a consistent image quality independent of view angle and object distance.

Reflective imaging systems are commonly used in astronomy due to their low weight and cost compared to lens-based systems of similar aperture diameter. An additional benefit of reflective systems is the lack of chromatic aberrations which makes them ideal for wide band imaging. However, reflective optics are typically restricted to field angles of not more than a few degrees without obscuring large parts of the aperture or without introducing prohibitively large off-axis aberrations. The construction of an efficient, compact, wide-field reflective system can bring the advantages of reflective systems to the wide class of applications where a larger field-of-view is demanded. Such a system could be achieved by mitigating the aberrations with a combination of pupil-phase modulation and digital image restoration.

Hybrid imaging techniques have also been used by Tremblay et al. in the design of a thin, four-reflection folded Cassegrainian imaging system to alleviate the effects of tolerance- and chromatically-induced defocus [124]. Although the design is in principle

reflective, chromatic aberrations are present due to the calcium fluoride medium between the mirrors that is required to increase the field-of-view to 17° without increasing the pupil obscuration beyond 81%. The combination of cubic pupil-phase modulation and image restoration was also used by Yan and Zhang to increase the defocus tolerance of an f/6 reflection-only system with a field-of-view of $1.5^\circ \times 5^\circ$ [126, 125]. Unfortunately the design of the system does not seem to be susceptible to adaptation for a significantly wider field of view.

In the next section the two-mirror Yolo reflector and its adaptation for wide-field imaging are introduced. Although large off-axis aberrations and obscurations are typical for wide-field reflective systems, the Yolo design is unobscured and has relatively modest off-axis aberrations. Section 4.3.2 describes the hybrid design process, and section 4.3.3 gives the results of the simulation and optimisation process, followed by a discussion of the holistic design approach.

4.3.1 Wide field reflective imaging

A Yolo reflector is an unobscured, tilted component telescope with two concave mirrors of comparable curvature and tilt [127, 128]. The design is similar to the Schiefspiegler but eliminates coma by an adequate choice of the mirror angles to the optical axis, while astigmatism is reduced by making the secondary mirror a toroidal surface. The design is useful for astronomy where large f-numbers and small field-of-view are practical. However, if the axial astigmatism is ignored, the field can be expanded to as large as $30^\circ \times 22.5^\circ$ because the light path is not obstructed. The adapted design shown in figure 4.10 has an aperture stop of 2 mm diameter at the primary mirror, and an effective focal length of 10 mm, considerably faster than typical reflective imaging systems. The high optical etendue and relatively small size make the design amenable for use in low-cost compact camera systems. The Yolo reflector has usually a comparable distance between the two mirrors (P - S) and the image plane (S - sensor); however, the mirrors of the adapted design are placed close together so that the image plane can be moved further down to avoid stray light reaching the sensor.

The off-axis aberrations of the Yolo design are modest considering the relatively wide field-of-view for a reflective system, and the design could already be useful as a low-cost foveal imaging system [129, 107]. However, for general purpose imaging, residual astigmatism severely limits the image quality outside the central region of the field-of-view. Fortunately the residual aberrations can be mitigated by using a combination of pupil modulation and image restoration. Pupil-phase modulation is introduced by modulation of the primary mirror surface with the purpose of obtaining an acceptable image across the full field-of-view and for object positions between 25cm and infinity. A third-order polynomial modulation $s(x,y) = \alpha(x^3+y^3)+\beta(x^2y+xy^2)$ in the normalised

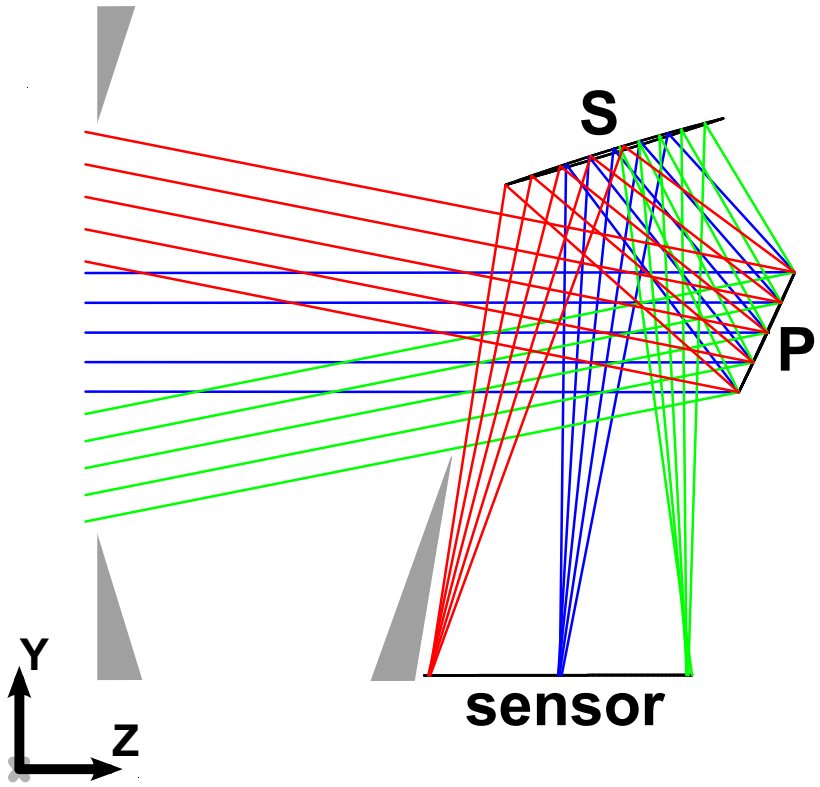


Figure 4.10: *The wide-field design of a two-mirror Yolo reflector. Light enters along the Z-axis and reflects of the primary spherical mirror (P) towards the secondary toroidal mirror (S) where it is focussed on the sensor array below. Gray areas indicate suggested light absorbing baffles.*

Cartesian pupil coordinates x and y, is chosen for its ability to mitigate defocus, and off-axis astigmatism in particular [94]. The parameters α and β of the surface-sag modulation are optimised together with the surface curvatures using the approach set out in the previous chapters.

4.3.2 Evaluation of hybrid imaging fidelity across the field-of-view

The design should produce a high fidelity image close to the geometric image for a field-of-view of $30^\circ \times 22.5^\circ$, and for object distances from 25 cm to infinity. To that end the point-spread function is calculated with high accuracy for 4×3 fields covering the object space of interest uniformly in field angle and at object distances of 25cm, 50cm and infinity.

It can be appreciated from figure 4.11 that the PSF changes only gradually with field angle. By assuming a linearly changing deconvolution kernel, spatially-variant image restoration algorithms can be implemented efficiently by deconvolution on overlapping image sections and recombination as proposed by Trussell and Hunt [130, 131].

Slower, however more generally applicable iterative algorithms, could also be used [132–134]. The change of the PSF with scene depth is however not known *a priori*, and in this section the spatially-variant deconvolution is therefore done with the PSFs at 50cm, independently of the actual object distance in the scene.

The metric combines the root-mean-square imaging errors of all sample-fields and object distances using the L^2-norm. By minimising this metric, the sum of the squared mean imaging error and its variance are minimised; guaranteeing consistent performance of the imaging system across its field-of-view and working range.

Figure 4.11: *Simulation of the detected image of the design without pupil-phase modulation.*

4.3.3 Results

Two hybrid designs are optimised for highest imaging fidelity across the field-of-view and through-focus from 25cm to infinity. The second design differs from the first only by the addition of a parameterisable third order modulation of the primary mirror surface at the aperture stop. The radius of curvature of both mirrors, and the radius of the toroid of the secondary mirror are optimised in the first design. The optimisation of the second design also includes the parameters of the surface-sag modulation. An optimal surface-sag modulation, $s(x,y) = 0.0942(x^3+y^3) - 0.896(x^2y+xy^2)\,\mu\text{m}$ was found, where x and y are normalised Cartesian pupil coordinates. The introduced phase modulation

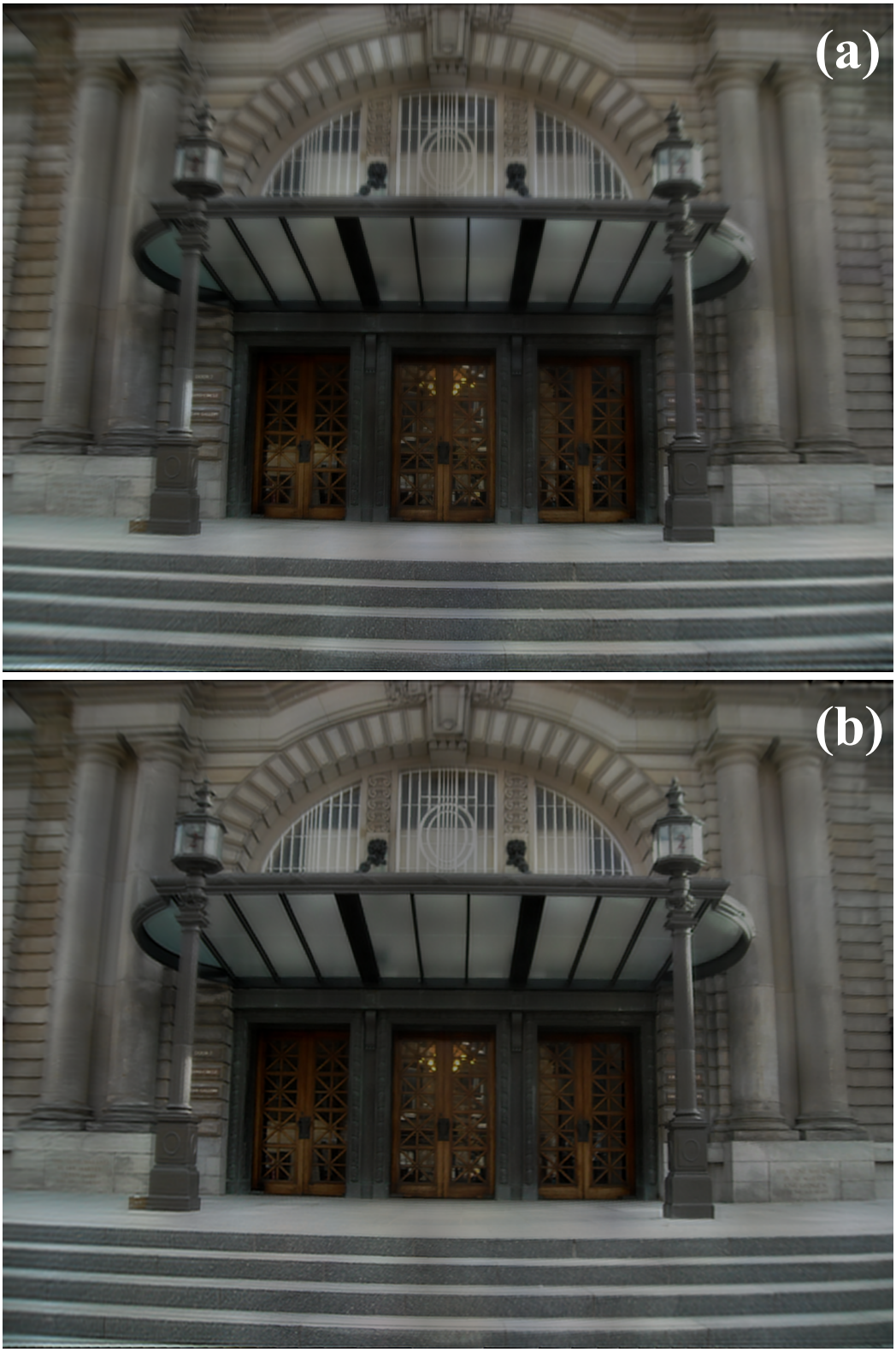


Figure 4.12: *Distortion-corrected and spatially-variantly deconvolved images of (a) the system without pupil modulation, and (b) the system with a generalised-cubic modulation of the primary mirror. Both images are simulated for scenes at infinity, while the kernel used in the deconvolution is calculated for the nominal object distance of 50cm. Enlarged sections are shown in figures 4.13 and 4.14.*

has $\beta \approx -10\alpha$, clearly distinct from earlier reported phase-modulations ($\beta = -3\alpha$, or $\beta \approx 0$) that minimise the imaging error under defocus. Although it has not yet been verified, this discrepancy could be due to the influence of off-axis aberrations, consisting mainly of astigmatism for most of the field-of-view.

Both designs are evaluated using image simulations carried out in *Zemax* using 25×25 diffractive PSF calculations and a projection on a 2000×1500 sensor array with a pixel pitch of $1.5\,\mu$m which introduces some aliasing. An in-focus test scene of 2000×1500 pixels is spatially-variantly convolved with out-of-focus PSFs at infinity for the wavelengths: 486.13 nm, 587.56 nm, and 656.27 nm, to simulate the respective colour channels. All PSFs are obtained using diffractive calculations on a 128×128 grid for both the pupil and the image. The simulation of the detected image is shown in figure 4.11 for the system without pupil-phase modulation.

Before further digital processing, the distortion seen in figure 4.11 is corrected by standard distortion-correcting interpolation. The smoothly changing PSF permitted efficient spatially-variant deconvolution by subdividing the image in rectangular sections, resulting in an execution time that is dominated by the computation of fast Fourier transforms performed on each section. Figure 4.12a shows the corrected image sharpened by deconvolution using 20×15 point-spread functions calculated across the field-of-view at the nominal focal distance of 50cm. A field-angle dependent blur can still be noticed; this is caused by off-axis aberrations that prevent accurate image restoration.

The image of the holistic design with pupil-phase modulation is corrected using the same procedure and shown in figure 4.12b. In contrast to the system without pupil-phase modulation, a more consistent image quality across the field-of-view can be noticed. Closer investigation of a region with modest aberrations near the optical axis, and one with considerable aberrations near the top-left lantern makes the difference between the two designs more obvious. Figures 4.13a and 4.13b show a magnification of the central area in respectively figures 4.12a and 4.12b.

Near the optical axis, the difference in image quality between both systems is modest, however details near the edge already show the improvement of the system with the pupil-phase modulation. The difference in image sharpness is most prominent at large field angles as can be seen from figure 4.14, showing a magnification of the region near the top-left lantern, approximately at 20° from the optical axis. Figure 4.14a and 4.14b show identical sections from figure 4.12a and 4.12b, corresponding respectively to the design before, and after the introduction of the phase-modulation.

4.3.4 Conclusions

A holistic approach to the design process can lead to the discovery of optical systems that only in conjunction with a digital post-processing step can meet preset specifications.

(a) **(b)**

Figure 4.13: *Enlargement of the central door in figure 4.12a and 4.12b, corresponding to: (a) the imaging system without pupil-phase modulation, and (b) the system with pupil-phase modulation.*

(a) **(b)**

Figure 4.14: *Detail of top left of the restored image corresponding to the design, a) without, and b) with modulation of the primary mirror.*

The integration of a hybrid optical-digital image quality metric into existing optical-design software enables the analysis and optimisation of such closely integrated systems.

This enables the conversion of a reflective design with limited field-of-view into a wide-field imaging system. Consistent imaging quality across the field-of-view could only be obtained with a combination of pupil modulation and digital image restoration. The holistic design approach allows the pupil modulation to accommodate the specificities of the considered design, leading to an optimal phase-modulation distinct to previously reported forms. This stresses the importance of a full integration of the hybrid optical-digital design process.

4.4 Experimental implementation of an extended depth-of-field imaging system

The simulation and optimisation software developed in the course of this thesis has been used for the design of an extended depth-of-field thermal imaging system. A prototype of the optimised design has been fabricated by Qioptiq Ltd., and compared to the equivalent imaging system designed using traditional optical design tools. Figure 4.15 shows images obtained for both systems and the subject at approximately 25 m (top), 60 m (centre), and 100 m (bottom row). Both, the traditionally designed and the hybrid designed imaging systems are optimised for a focal distance of 20 m. Notice that although the sharpness for both systems is comparable near the focal plane, at larger distances the image sharpness deteriorates quickly for the traditionally designed system. On the other hand, the reduction in sharpness for the hybrid imaging system is noticeably less. This is most apparent at large defocus as can be seen by comparing the triangular traffic sign at the end of the road for both imaging systems.

The system is approximately f/1, and incorporates the trefoil phase-modulation in the back surface of the first lens element. This phase-modulation is optimised so that the hybrid design yields the highest imaging fidelity using the method described in this thesis.

The image is recorded with an uncooled bolometric detector with 35 µm pixel pitch, resulting in an under-sampling factor of approximately four. To facilitate electronic implementation, the 320×240 pixel recorded image is deconvolved in the spatial domain using a linear filter kernel of size 27×27 pixels.

4.5 Conclusion

The integration of commercial off-the-shelf optical-design software with general-purpose computing software has been shown to be practical for the design of hybrid optical-digital imaging systems. This is demonstrated with the optimisation of a defocus-tolerant

Figure 4.15: *Images from a video sequence captured with two equivalent thermal imaging systems; left: without pupil modulation, right: optimised for defocus tolerance by incorporating a trefoil modulation at the back surface of the first lens element. The contrast between the road markings and the tarmac is approximately* $2\,K$.

thermal imaging system, and compared to the optimisation of a Nyquist-sampled paraxial system. Although the metric calculation requires more time than that of conventional metrics, an optimum is typically found on a standard desktop computer within one hour. With further improvements in computing hardware and the algorithms, the calculation time is expected to become of less importance.

The most common radially-symmetric and antisymmetric phase modulations have

been compared for use in two hybrid imaging systems employing a simple Wiener filter deconvolution without *a priori* knowledge of the amount of defocus, and a conservative peak-signal-to-noise ratio of 100. The optimal magnitude of most phase modulations was found to be similar to that when using the optimal image restoration algorithm assumed in chapter 2; however, the optimal magnitude of the phase-modulation corresponding to trefoil is found to be more than twice as large when incorporated into a Nyquist-sampled paraxial system. It was found that when employing a simple Wiener deconvolution, the radially-symmetric phase modulations yield a lower imaging error than the trefoil modulation.

The logarithmic axicon and the logarithmic asphere appear to converge to the quartic phase modulation which was found to yield the lowest expected imaging error of the radially-symmetric phase-modulations. Furthermore, the quartic phase modulations is more practical since the associated cost function is less-irregular than that of the phase functions that contain a logarithm. Instead of evaluating the different radially-symmetric modulation types, one could optimise only the quartic, perhaps extended with higher order terms since these have been shown to further reduce the expected imaging error by a modest amount.

With the reflective optics example it was shown that hybrid optical-digital imaging systems need not be restricted to extension of the depth-of-field. Also off-axis aberrations can be accounted for by optimisation of the hybrid imaging fidelity across the full field-of-view. The optimisation of the curvatures and the surface-sag modulation of the primary mirror enabled the design of a compact wide-field reflective imaging system. The optimal modulation was atypical for defocus compensation, and appears to aid in the correction of residual off-axis astigmatism.

Chapter 5 – Aberration mitigation using amplitude modulation

5.1 Introduction

So far, only pupil-phase modulation has been considered for the mitigation of aberrations. In this chapter the hybrid imaging quality metric introduced in chapter 2 is demonstrated for the optimisation of amplitude-modulated adaptive optics. Although the pupil phase can be corrected using active or adaptive optics [135–137], wave-front modulators tend to be high cost and the necessary reflection geometry can be a disadvantage. Low-cost, transmissive, liquid-crystal phase-modulation on the other hand is limited to use at visible and near-infrared wavelengths. Binary phase-modulation using ferroelectric liquid crystals has been considered as an alternative to continuous modulation [138]; however binary-amplitude modulation offers lower complexity and cost, and can be accomplished efficiently from ultra-violet to far-infrared wavelengths with a wide variety of fixed or agile spatial-light amplitude-modulation techniques [76–79]. It will be shown that a binary-amplitude mask located at the aperture stop can be optimised to mitigate phase aberrations and allow sufficient information to be recorded for the recovery of a sharp image using standard digital-image restoration.

In this chapter an upper limit is derived for the asymptotic MTF obtainable in the presence of large aberrations, using amplitude masks in general. Although a mask could be devised to yield the asymptotic MTF for specific spatial frequencies, it is unlikely that masks exist that approach the limit at all spatial frequencies simultaneously. However, significant image contrast across all spatial frequencies is obtained for masks that selectively block destructively interfering parts of the aperture and such a mask can be considered to be the amplitude mask equivalent of the phase mask proposed by Love et al. [139, 140]. Here an analytical expression is derived for the MTF when masking arbitrarily large aberrations, which can be seen to be in agreement with the Strehl ratio found for the phase mask in [139, 141]. Furthermore, in this chapter it is shown that for aberrations as large as 3.5 waves of root-mean-square optical-path difference, part of the MTF can be increased beyond the asymptotic limit by optimisation of three additional free parameters of the mask. More recently, a similar binary-amplitude modulation has been suggested for the correction of residual aberrations in adaptive optics by Osborn et al. [142]. Rigourous simulations showed that the contrast can be improved significantly by selectively blocking areas of the pupil where the wave-front aberration surpassed a chosen threshold. Sufficient mask transmission is maintained because the residual aberrations are relatively modest. In contrast, the technique described here enables high optical transmission for arbitrarily high aberrations.

In the following section, the asymptotic MTF is derived for large aberrations, and for amplitude masks in general. In section 5.3, contour masks are introduced and an

analytical expression is derived for their MTF. Performance in the presence of common aberrations is evaluated for paraxial imaging with monochromatic light, and is found to be in close agreement with the derived MTF; however, some improvement over the asymptotic case was obtained by optimisation of the free parameters of the mask. Finally, the contour mask is evaluated for off-axis imaging with a compound lens in broadband light. Although the contrast is lower than that predicted for monochromatic illumination, the resultant contrast is deemed sufficient to enable digital image recovery for this lens.

5.2 Contrast obtainable in the presence of large aberrations

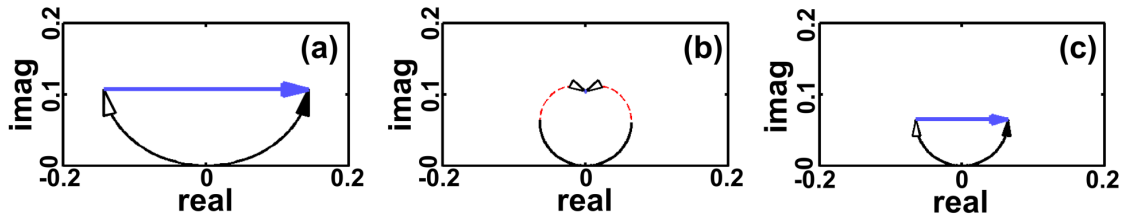


Figure 5.1: *Decompositions of the OTF integral in the complex plane at spatial-frequency* $\boldsymbol{\nu}$. *The black and dashed red curves depict respectively, constructive and destructive contributions* $\mathrm{d}\mathbf{u}$ *to the net OTF (blue phasor), for defocus of a circular pupil for (a)* $W_{20} = \lambda/4$, *(b)* $W_{20} = 0.642\lambda$ *(for which a null occurs), and (c), with a putative mask blocking the destructive interference components yielding a positive OTF.*

The influence of an aberration on the OTF of an optical system can be lucidly understood from the autocorrelation of the pupil function [45]:

$$OTF(\boldsymbol{\nu}) \quad = \quad \frac{1}{\mathrm{supp}\,(P)} \iint P^{*}(\mathbf{u}-\boldsymbol{\nu})\,P(\mathbf{u}+\boldsymbol{\nu})\,\mathrm{d}\mathbf{u}, \tag{5.1}$$

with the pupil function, $P(\mathbf{u}) = \exp(\mathrm{i}\,l(\mathbf{u}))$, $\forall \mathbf{u}\ \|\mathbf{u}\| \leqslant 1$, and 0 otherwise; where $l(\mathbf{u})$ is the optical path length variation expressed in radians; while $\mathbf{u}$ and $\boldsymbol{\nu}$ are the normalised pupil coordinate and spatial-frequency respectively. The OTF is normalised by division by $\mathrm{supp}\,(P)$, the support of the pupil function, equal to π for a circular aperture. For our purpose equation (5.1) can be simplified to:

$$OTF(\boldsymbol{\nu}) \quad = \quad \frac{1}{\mathrm{supp}\,(P)} \iint_{\Omega(\boldsymbol{\nu})} \exp(\mathrm{i}\,\Delta(\mathbf{u},\boldsymbol{\nu}))\,\mathrm{d}\mathbf{u}, \tag{5.2}$$

where $\Delta(\mathbf{u},\boldsymbol{\nu}) = \varphi(\mathbf{u}+\boldsymbol{\nu}) - \varphi(\mathbf{u}-\boldsymbol{\nu})$, is the difference in optical path length between two points on the pupil $2\boldsymbol{\nu}$ apart; and $\Omega(\boldsymbol{\nu})$ is the integration area of the autocorrelated pupil.

It is informative to decompose the integrand in equation (5.2) by plotting definite integrals with increasing limit in the complex plane [99], as depicted in figure 5.1 for a circular aperture for $\|\boldsymbol{\nu}\| = 1/2$. In the case of a modest defocus $W_{20} = \pi/4$, then $-\pi/2 \leqslant \Delta(\mathbf{u},\boldsymbol{\nu}) \leqslant \pi/2$, as can be seen from the slope of the curve in figure 5.1(a), and all components contribute constructively to the resultant phasor. For larger aberrations, destructive interference contributes to the resultant phasor and leads to a reduced MTF, or nulls as shown in figure 5.1(b). By devising an appropriate pupil mask, excessive suppression of the MTF can be avoided by removing or reducing those destructive contributions, as depicted in figure 5.1(c). It will be shown that, in the presence of large aberrations, sufficiently high contrast can be guaranteed to enable the digital recovery of a high-quality image.

An upper limit for the MTF with amplitude mask $M(\mathbf{u}) \in [0,1]$, can be calculated for large aberrations for which $\Delta(\mathbf{u},\boldsymbol{\nu}) \bmod 2\pi$ in equation 5.2 can be considered to be uniformly distributed in the half-open[1] interval $[0,2\pi)$. The ideal mask would permit only constructive interference for constituent image phasors, i.e. for $(\Delta(\mathbf{u},\boldsymbol{\nu}) + \pi/2) \bmod 2\pi \leqslant \pi$; and block all destructive interference. The to-be-minimised or maximised phasor magnitude depends on the product of the mask value at two points in the pupil, hence partial pupil-attenuation can never achieve better contrast than a binary mask. Without loss of generality, the discussion is therefore restricted to binary masks: $M(\mathbf{u}) \in \{0,1\}$. Assuming that the ideal mask exists, in general for $\|\boldsymbol{\nu}\| > 0$ the double integral in equation (5.2) can be written as a single-dimensional Lebesgue integral [143], and approximated as shown in section 5.4.1 for large aberrations by:

$$OTF_i(\boldsymbol{\nu}) = \frac{1}{\operatorname{supp}(P)}\frac{1}{T}\int_{-\pi/2}^{\pi/2} \exp(\mathrm{i}\,\Delta)\mu'(\boldsymbol{\nu},\Delta)\mathrm{d}\Delta \approx \frac{1}{\operatorname{supp}(P)}\frac{1}{T}\frac{A_\Omega(\boldsymbol{\nu})}{2\pi}\int_{-\pi/2}^{\pi/2} \exp(\mathrm{i}\,\Delta)\mathrm{d}\Delta \quad (5.3)$$

where, in the limit for large aberrations, $\mu'(\boldsymbol{\nu},\Delta) \to A_\Omega(\boldsymbol{\nu})/2\pi$ accounts for the change in integration variables, and $1/T$ renormalises the transfer function for a mask with transmission, T. The area of overlap of the autocorrelated pupil function, $A_\Omega(\boldsymbol{\nu})$, at spatial-frequency $\boldsymbol{\nu}$, is equal to the product of the area of support of the pupil function, $\operatorname{supp}(P)$ and the diffraction-limited MTF at $\boldsymbol{\nu}$, $MTF_{\mathrm{DL}}(\boldsymbol{\nu})$. Considering also that $T \to 1/2$ for large aberrations, the upper limit of the MTF with the putative amplitude mask is given by:

$$MTF_{\mathrm{p}}(\boldsymbol{\nu}) \approx \frac{2}{\pi} MTF_{\mathrm{DL}}(\boldsymbol{\nu}) \approx 0.64 MTF_{\mathrm{DL}}(\boldsymbol{\nu}). \quad (5.4)$$

For a specific spatial-frequency $\|\boldsymbol{\nu}\| \geqslant 1/2$ it is straightforward to find a mask satisfying equation (5.4); however, for a broad range of spatial frequencies and an arbitrary aberra-

[1] The probability of obtaining a phase of 2π after the modulus operation is zero, so strictly speaking $\Delta(\mathbf{u},\boldsymbol{\nu}) \bmod 2\pi$ can only have non-zero values in the half-open interval $[0,2\pi)$.

tion such a mask might not exist. The optimal mask can be found by a rigourous search over a high-dimensional discretised mask space as proposed by Stayman et al. [144]. Although such an approach is practical when the aberration and optimal mask are fixed, for dynamic aberrations requiring adaptive optimisation of masks, the computational burden for this approach is probably too great for most real-time applications. Some of the masks reported by Stayman et al. show remarkable regularity. To investigate this further, the global optimisation process was repeated for this thesis, and it was found that after sufficient iterations, the mask obtained for a defocus of $W_{20} = \lambda$ closely resembles the discretised Fresnel zone plate. Figure 5.2 shows a 16×16 and a 32×32 mask after respectively 1000 and 2000 iterations of the global optimisation differential-evolution algorithm [117]. This algorithm generalises the genetic algorithm by replacing one of the crossover vectors with a linear combination, defined by the differential weight, of three other vectors. All vectors are picked randomly from a population of size 50, the probability of crossover is 80%, and the differential weight or step size is chosen to be 0.9. The masks seem to tend towards blocking contours of equal phase. Note that the

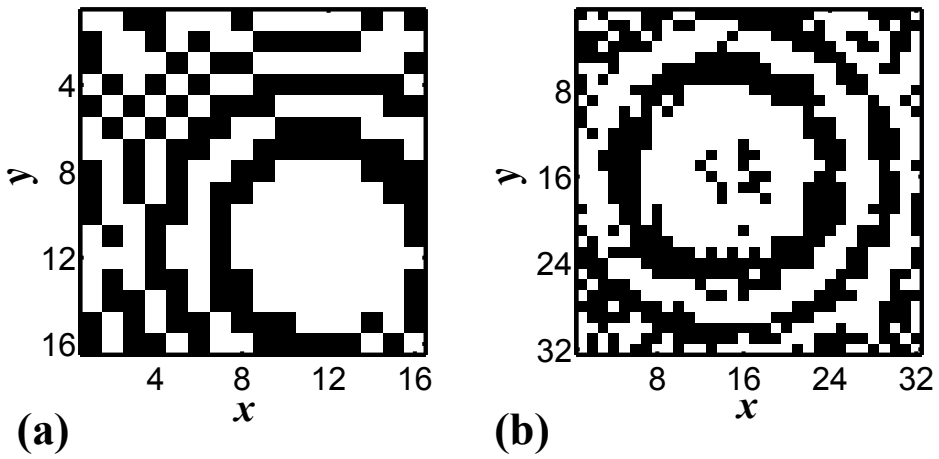


Figure 5.2: *Two discretised binary amplitude masks optimised for the mitigation of pure Zernike defocus $W_{20} = 2\lambda$; (a) the optimal mask found for a 16×16 mask after 1000 iterations of the global optimisation algorithm, and (b) the optimum found for a 32×32 mask after 2000 iterations. Transmissive areas are indicated in white, non-transmissive in black.*

centre of the rings of the masks in figure 5.2 is different for both simulations. Displacing a Fresnel zone plate merely shifts the image, there is therefore no reason why the centre of the rings should coincide with the centre of the mask. In fact, since the pupil is finite and since the mask is restricted to one of the $2^{16\times16}$ masks, it could be that the sharpest image is produced off-axis.

Even if the discretised masks converge to block areas of equal phase, three degrees of freedom remain: wavefront tip, tilt, and piston; the tip and tilt when image shifts are not considered important, and piston because an arbitrary reference phase can be chosen

for transmission. However, in contrast to the high-dimensional space of discretised masks, it is now possible to confine the optimisation process to the more tractable three dimensional space of the *contour* masks, formally defined in the next section.

5.3 Contour Masks

The use of contour masks is proposed here, since they yield comparable contrast to the discretised masks while only requiring low-dimensional optimisation. The contour mask, $M(\mathbf{u})$, for an aberration characterised by an optical path length variation, $l(\mathbf{u})$, is defined as:

$$M(\mathbf{u}) = 1, \forall \mathbf{u} \left| \left(l(\mathbf{u}) - \varphi_0 - \mathbf{t}\cdot\mathbf{u} + \frac{\Delta\varphi}{2} \right) \mod 2\pi \leqslant \Delta\varphi \right. \text{, and 0 otherwise,} \quad (5.5)$$

where $\Delta\varphi$ is the maximum permitted phase-difference between any two points in the pupil. The to-be-optimised parameters φ_0 and $\mathbf{t}$ are the reference phase and the combination of tip and tilt respectively. In the simple case of defocus, $M(\mathbf{u})$ is a Fresnel zone plate. The introduction of tip and tilt with the dot product, $\mathbf{t}\cdot\mathbf{u}$, merely displaced the imaged field-of-view. For modest displacements such imaging systems can be considered equivalent, hence tip and tilt are included as an additional free parameter in the optimisation process, potentially improving the contrast further.

5.3.1 Contrast for monochromatic illumination

More generally, and for sufficiently large aberrations the pupil phase, $\varphi(\mathbf{u}) = l(\mathbf{u}) \mod 2\pi$ can be considered to be uniformly distributed in the interval $[0, 2\pi)$, and for $\|\boldsymbol{\nu}\| > 0$, the phases $\varphi_1 = \varphi(\mathbf{u} - \boldsymbol{\nu})$ and $\varphi_2 = \varphi(\mathbf{u} + \boldsymbol{\nu})$ are independent, so that the integral in equation (5.2) can be approximated as follows:

$$OTF(\boldsymbol{\nu}) \approx \frac{MTF_{\mathrm{DL}}(\boldsymbol{\nu})}{4\pi^2} \int_{-\pi}^{\pi} \exp(-i\varphi_1) \int_{-\pi}^{\pi} \exp(i\varphi_2) \mathrm{d}\varphi_2 \mathrm{d}\varphi_1, \quad (5.6)$$

where the factor $MTF_{\mathrm{DL}}(\boldsymbol{\nu})/4\pi^2$ accounts for the change in integration variables as demonstrated in section 5.4.2. The integrals evaluate to zero as expected for large aberrations. Contour masks on the other hand permit only interference of phases $-\Delta\varphi/2 \leqslant l(\mathbf{u}) \mod 2\pi \leqslant \Delta\varphi/2$, so that the OTF can be calculated by changing the integration limits in equation (5.6) to give a contour mask OTF without nulls:

$$OTF_c(\boldsymbol{\nu}) \approx \frac{2\pi}{\Delta\varphi} \frac{MTF_{\mathrm{DL}}(\boldsymbol{\nu})}{\pi^2} \sin^2\left(\frac{\Delta\varphi}{2}\right) \leqslant \frac{2}{\pi^2} MTF_{\mathrm{DL}}(\boldsymbol{\nu}) \quad (5.7)$$

It can be seen that before renormalization[2] by $2\pi/\Delta\varphi$, the contrast is maximised for $\Delta\varphi = \pi$ when, in the limit for large aberrations, $T = 50\%$. Following from equation (5.7), such a mask will yield approximately 20% $(2/\pi^2)$ of the contrast of the diffraction-limited transfer function. More importantly, the expected MTF is independent of the magnitude of the aberrations: it is limited only by the contrast and spatial resolution of the spatial-light modulator.

The same analysis can be used to derive the OTF for the binary phase masks proposed by Love et al. [139]. It can be seen that this leads to an OTF equal to $4/\pi^2$ times that of the diffraction-limited MTF. Since this OTF is real and non-negative, the maximum intensity in the image plane can be found on the optical axis. The fraction of the integral of the OTF and the integral of the diffraction-limited MTF, yields a Strehl ratio of $4/\pi^2$, in agreement with that found in [139].

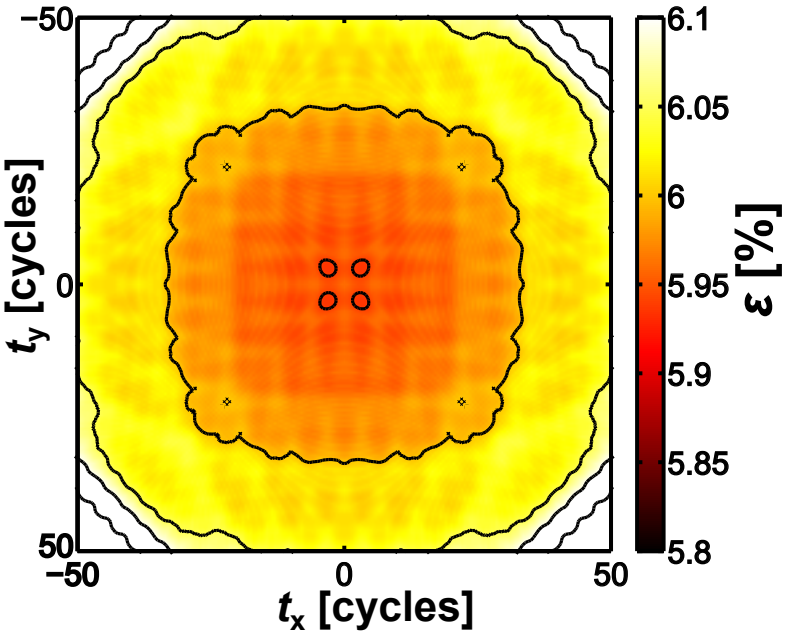


Figure 5.3: *Typical cost-function shown for astigmatism as a function of tip and tilt, (t_x, t_y), and minimised for phase-reference, ϕ_0. Bright yellow and white regions indicate a high cost for large tip or tilt. The encircled dark red spots in the centre indicate four modest tip and tilts that lead to equivalently performing masks, one of which is shown in figure 5.4a.*

Only the parameters φ_0 and $\mathbf{t}$ are varied to optimise the MTF, or more specifically, the expected imaging error as defined in chapter 2 and [114]. As can be seen from the contour plot in figure 5.3, this low-dimensional optimisation space contains local minima; however a minimum very close to the global minimum is typically achieved after only a few iterations of the differential-evolution algorithm [117] using a population

[2]The MTF can be further increased at the cost of reducing the transmission T. With renormalisation, it is easy to verify that the MTF is maximised when $\tan(\Delta\varphi/2) = \Delta\varphi$, for $\Delta\varphi \in (0, \pi)$. This is satisfied for $\varphi = 2.33112\ldots \approx \frac{3\pi}{4}$, with corresponding transmission $T \approx 37.1\%$.

of 30, a crossover rate of 0.8, and a differential-evolution step size of 0.9.

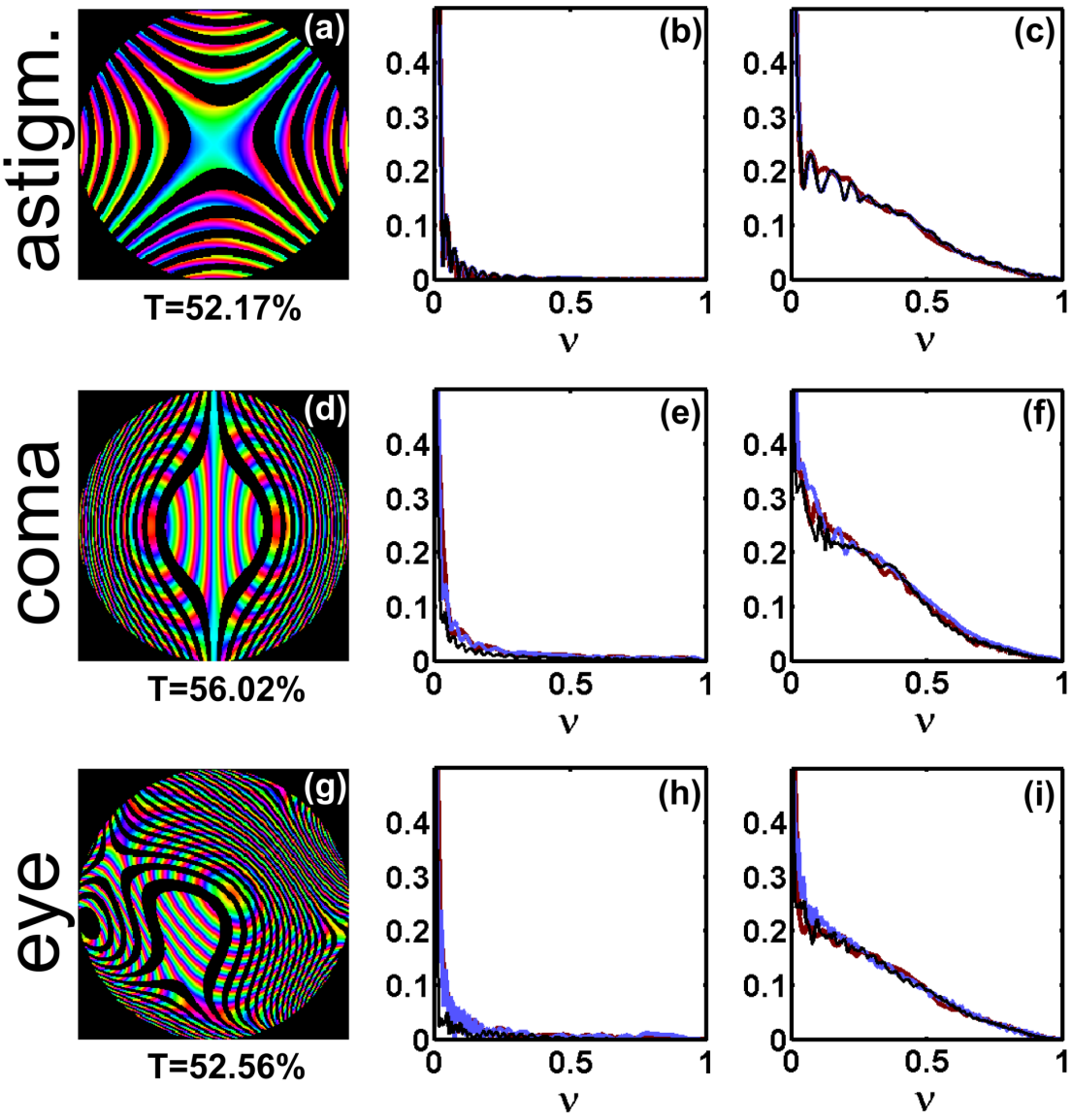


Figure 5.4: *Aberration compensation of astigmatism (a-c), coma (d-f), and an aberration representative for the human eye (g-i), of respectively a root-mean-square optical-path difference of* 2λ*,* 2λ*, and* 3.5λ*. The aberration phases and contour masks with the relative transmission noted below are shown in (a,d,g). The tangential (black), sagittal (blue), and diagonal (red) MTF without mask in (b,e,h), and with mask in (c,f,i).*

The three masks depicted in the left column of figure 5.4 were optimised for minimum image error for three representative aberrations: astigmatism, coma and an irregular phase aberration not atypical of the human eye [10, 145, 12]. Black areas represent opaque areas of the mask, and the hue of the transmissive areas indicates pupil phase. The plots in the central column of figure 5.4 show the MTFs for the non-masked, aberrated pupils: the low values and nulls in the sagittal, diagonal, and tangential MTF are readily apparent. The MTFs in the right column correspond to the masked, aberrated

pupils and show a significant increase in the MTFs. Even considering that the total transmitted intensity is reduced by the mask to approximately 50%, the MTF for the masked pupil is significantly increased for all spatial frequencies and there are no nulls, thus enabling the recovery of a high quality image to be recovered by digital processing.

Note that the transmission of the masks is actually larger than 50%. Although the tested aberrations are considerable, areas of linearly changing phase exist so that the contour mask can have large open areas such as seen at the centre of figure 5.4d. Furthermore, when at low spatial frequencies this open area coincides with itself in the autocorrelation of the pupil function, it evokes a modest boost in the MTF as can be seen in figure 5.4f for spatial frequencies of $T \leq 1/2$.

5.3.2 *Contrast for a compound lens and broadband illumination*

Many applications are essentially monochromatic; for instance laser imaging, fluorescence microscopy, and retinal imaging with a scanning laser ophthalmoscope. Other applications have a modest bandwidth; e.g. true-colour imaging (relative spectral width $\Delta\lambda/\lambda_0 \approx 20\%$ per channel), snapshot multi-band imaging [146] ($\approx 2\%$), and long-wavelength infrared imaging with quantum-well detectors ($\approx 5\%$). In these cases one might expect that mask optimisation based on the above principle of monochromatic imaging might suffer some reduction in optical efficiency. As a practical example, a contour mask for enhanced off-axis imaging with a relative bandwidth of 20% is investigated now.

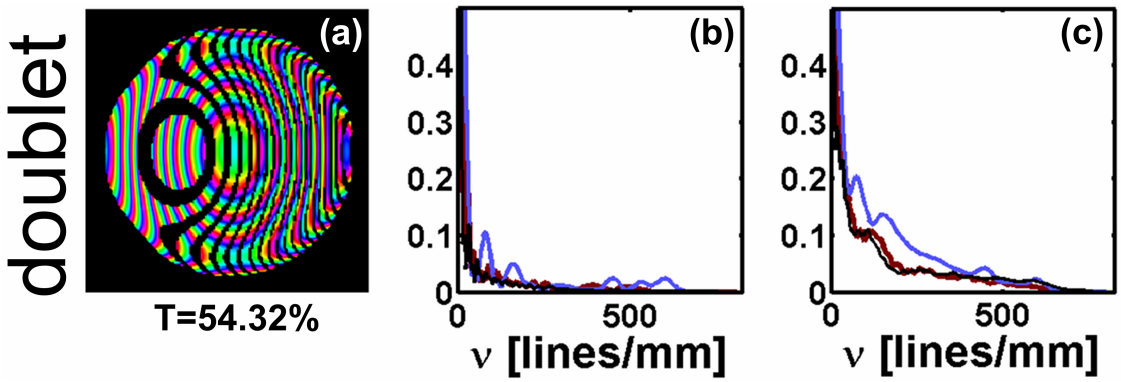


Figure 5.5: *(a) Off-axis aberration at* $\lambda_0 = 550\text{nm}$ *of a cemented doublet with contour mask optimised for panchromatic operation for* $495\text{nm} \leq \lambda \leq 605\text{nm}$ *(*$\Delta\lambda/\lambda_0 = 20\%$*). (b) The tangential (black), sagittal (blue), and diagonal MTF (red) before, and (c) after the introduction of the contour mask.*

A contour mask was optimised for correction of aberrations introduced by an f/5 cemented doublet when used 5° off-axis at a nominal central wavelength of $\lambda_0 = 550\,\text{nm}$. In this case the aberration consists mainly of astigmatism and field curvature with a combined peak-to-valley optical-path difference of approximately $10\lambda_0$. The

aberration and optimised mask are shown in figure 5.5a. MTFs calculated by ray tracing with broadband light (with wavelengths $495\,\text{nm} < \lambda_0 < 605\,\text{nm}$, uniformly weighted) are shown in figure 5.5b and figure 5.5c respectively before and after the introduction of the mask. It can be seen from figure 5.5b that without a mask the MTF is strongly suppressed and contains a large number of nulls. In contrast, as shown in figure 5.5c, the addition of the mask yields a higher MTF and an absence of nulls. Although the contour mask was designed for monochromatic operation, useful aberration correction properties are maintained for a compound lens illumination with an extended bandwidth; however, the contrast is lower than would be obtained for monochromatic operation; and although large aberrations can be corrected, the MTF will remain high only if the phase-aberration function is relatively invariant with wavelength.

5.4 Optical transfer function approximation for large aberrations

5.4.1 Single integral over the phase difference

The OTF defined as using the double Riemann integral of equation (5.2), can also be written using a Lebesgue integration [143], where one integrates over the possible function values $\hat{\Delta}$, of $\Delta(\mathbf{u},\boldsymbol{\nu})$ as follows:

$$OTF(\boldsymbol{\nu}) \;=\; \frac{1}{\operatorname{supp}(P)} \iint\limits_{\Omega(\boldsymbol{\nu})} \exp\left(\mathrm{i}\,\Delta(\mathbf{u},\boldsymbol{\nu})\right) d\mathbf{u} = \frac{1}{\operatorname{supp}(P)} \int\limits_{-\infty}^{\infty} \exp\left(\mathrm{i}\,\hat{\Delta}\right) \mathrm{d}\mu(\hat{\Delta}), \quad (5.8)$$

with $\mathrm{d}\mu(\hat{\Delta})$, the differential area of the integration region $\Omega(\boldsymbol{\nu})$ where $\hat{\Delta} \leqslant \Delta(\mathbf{u},\boldsymbol{\nu}) < \hat{\Delta} + \mathrm{d}\hat{\Delta}$.

Figure 5.6b shows a contour plot of the two-dimension Riemann integral; the hue indicates the complex argument of the integrand, $\exp\left(\mathrm{i}\,\Delta(\mathbf{u},\boldsymbol{\nu})\right)$, and grey contour lines indicate where $\Delta(\mathbf{u},\boldsymbol{\nu}) \mod 2\pi = \pi$. The Lebesgue integration can be seen as a one-dimensional integration of the probability density function (PDF) of the values $\Delta(\mathbf{u},\boldsymbol{\nu})/2\pi$ shown in figure 5.6a.

Since $\exp\left(\mathrm{i}\,\hat{\Delta}\right)$ is periodic with period 2π, the single integral can be split into a summation of a number of intervals of size 2π covering the image of $\Delta(\mathbf{u},\boldsymbol{\nu})$:

$$\int\limits_{-\infty}^{\infty} \exp\left(\mathrm{i}\,\hat{\Delta}\right) \mathrm{d}\mu(\hat{\Delta}) = \int\limits_{-\pi}^{\pi} \exp\left(\mathrm{i}\,\hat{\Delta}\right) \sum_{k=k_{\min}}^{k_{\max}} \mathrm{d}\mu(\hat{\Delta} + 2\pi k), \quad (5.9)$$

where the values $k_{\min}$ and $k_{\max}$ bound the image of $\Delta(\mathbf{u},\boldsymbol{\nu})$. In other words, instead of integrating over $\hat{\Delta}$, one can integrate over $\hat{\Delta} \mod 2\pi$, weighting the integrand with the total area of equally spaced contours, i.e. where $\hat{\Delta} \leqslant \Delta(\mathbf{u},\boldsymbol{\nu}) \mod 2\pi < \hat{\Delta} + \mathrm{d}\hat{\Delta}$. This

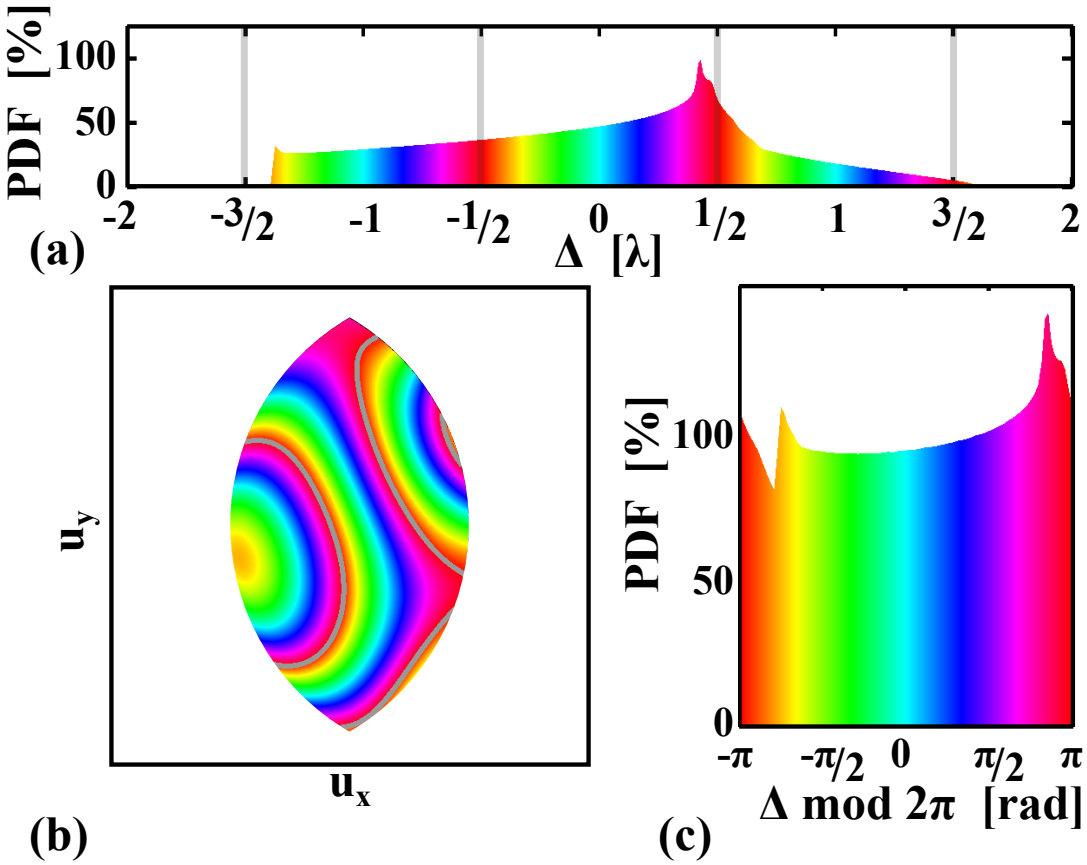


Figure 5.6: *Three decompositions of the OTF integral argument for a modest aberration with a peak-to-valley optical-path difference of* $\approx 1.5\lambda$. *The argument of the autocorrelation integral is directly dependent on the difference in optical path length between two points on the pupil,* $\Delta(\mathbf{u},\boldsymbol{\nu})$. *The function* $\Delta(\mathbf{u},\boldsymbol{\nu})$ *is shown in (b) for* $\boldsymbol{\nu} = (1/2, 0)$. *Lines of equal complex argument are indicated in grey, the complex argument of the integrand is indicated by the hue and its values can be read from (b). Plots (a) and (c) show the probability density of the values* $\Delta(\mathbf{u},\boldsymbol{\nu})/2\pi$, *and* $\Delta(\mathbf{u},\boldsymbol{\nu}) \mod 2\pi$ *respectively.*

is equivalent to integrating over the probability density function of the $\Delta(\mathbf{u},\boldsymbol{\nu}) \mod 2\pi$ shown in figure 5.6c.

Consider now a proportionally larger aberration, with $l_c(\mathbf{u}) = c \cdot l(\mathbf{u})$, so that the difference in optical path length is correspondingly larger: $\Delta_c(\mathbf{u},\boldsymbol{\nu}) = c \cdot \Delta(\mathbf{u},\boldsymbol{\nu})$ as that shown in figure 5.7, and its area of integration $\mu_c(\hat{\Delta})$, is related as: $\mu_c(\hat{\Delta}) = \mu(\hat{\Delta}/c)$. The summation inside the integral of equation (5.9) for the larger aberration can then be written as:

$$\sum_{k=c\cdot k_{\min}}^{c\cdot k_{\max}} \mathrm{d}\mu_c\left(\hat{\Delta}+2\pi k\right) = \sum_{k=c\cdot k_{\min}}^{c\cdot k_{\max}} \mathrm{d}\mu\left(\hat{\Delta}/c+2\pi k/c\right) 1/c \tag{5.10}$$

$$= \sum_{k=2\pi k_{\min}/\delta}^{2\pi k_{\max}/\delta} \mathrm{d}\mu\left(\hat{\Delta}\delta/2\pi+k\delta\right)\delta/2\pi, \tag{5.11}$$

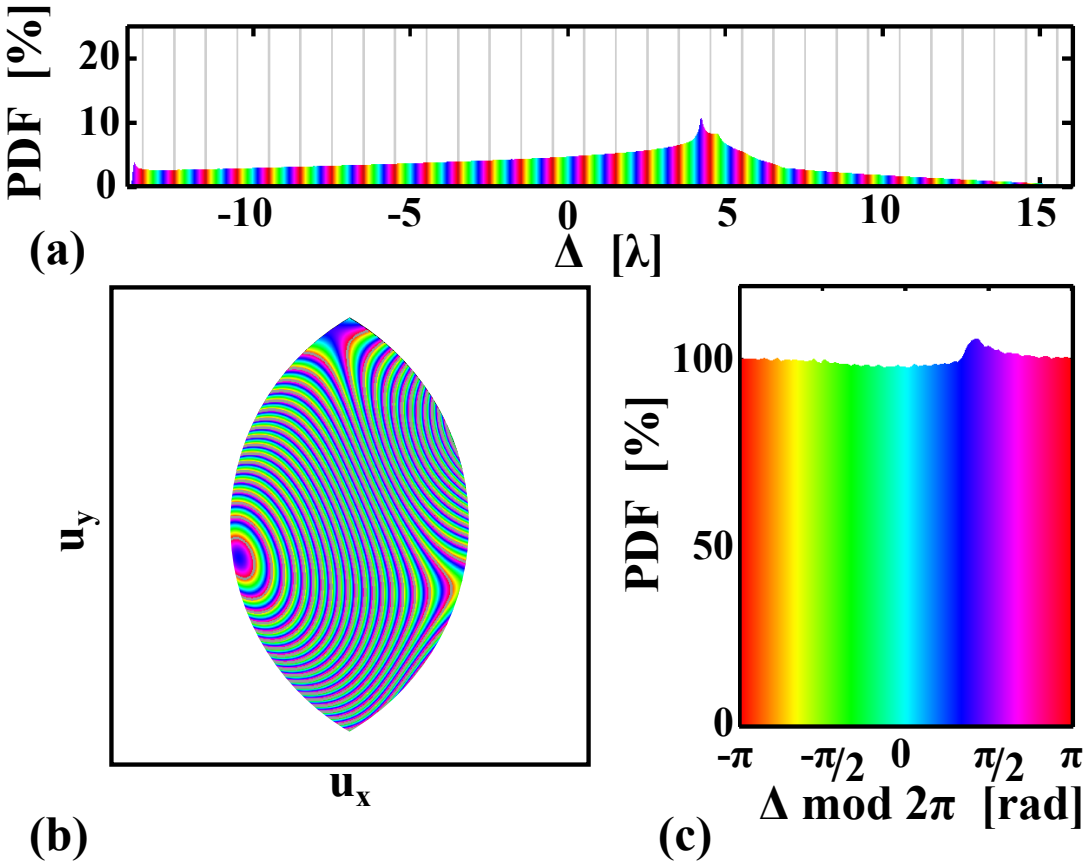


Figure 5.7: *The same three decompositions of the OTF integral for an aberration with ten times larger peak-to-valley optical-path difference of* $\approx 15\lambda$*. The complex argument of a large aberration is more uniformly distributed as can be seen by comparing (c) to the modest aberration shown in figure 5.6c.*

with $\delta = 2\pi/c$. Since $\hat{\Delta}$ is bound to $[-\pi, \pi]$, the term $\hat{\Delta}\delta/2\pi$ must vanish in the limit for $c \to \infty$. Graphically this can be understood as the convergence to a uniform probability density distribution in $[-\pi, \pi)$ seen in figure 5.7c. Hence, if the Lebesgue measure $\mu\left(\hat{\Delta}\right)$ is differentiable, equation (5.10) will converge to the integral:

$$\lim_{c\to\infty} \sum_{k=c\cdot k_{\min}}^{c\cdot k_{\max}} \mathrm{d}\mu_c\left(\hat{\Delta}+2\pi k\right) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{d}\mu\left(\hat{\Delta}\right)\mathrm{d}\hat{\Delta} = \frac{A_\Omega(\boldsymbol{\nu})}{2\pi}\mathrm{d}\hat{\Delta}, \tag{5.12}$$

with $A_\Omega(\boldsymbol{\nu})$, the total area of the integration region. The requirement of differentiability of $\mu\left(\hat{\Delta}\right)$ is not very restrictive. For the limiting cases where $\mu\left(\hat{\Delta}\right)$ is not differentiable, such as for a constant $\Delta(\mathbf{u}, \boldsymbol{\nu})$, the upper limit of equation (5.7) can be surpassed. E.g. for $\boldsymbol{\nu} = \mathbf{0}$, the OTF is unity. In general however, for large aberrations, the phase difference $\Delta(\mathbf{u}, \boldsymbol{\nu}) \mod 2\pi$ can be considered to be uniformly distributed, and the summation in equation (5.9) can be brought outside the integral so that equation (5.3) is obtained after substitution in equation (5.8).

5.4.2 Double integral over pupil phases of the autocorrelation

In general, for large aberrations, the phases $\varphi_1(\mathbf{u}) = l(\mathbf{u}-\boldsymbol{\nu}) \mod 2\pi$ and $\varphi_2(\mathbf{u}) = l(\mathbf{u}+\boldsymbol{\nu}) \mod 2\pi$ will also be approximately uniformly distributed. This is demonstrated graphically in figures 5.8 and 5.9, where figures 5.8b and 5.9b now show the optical-path difference introduced by the aberration.

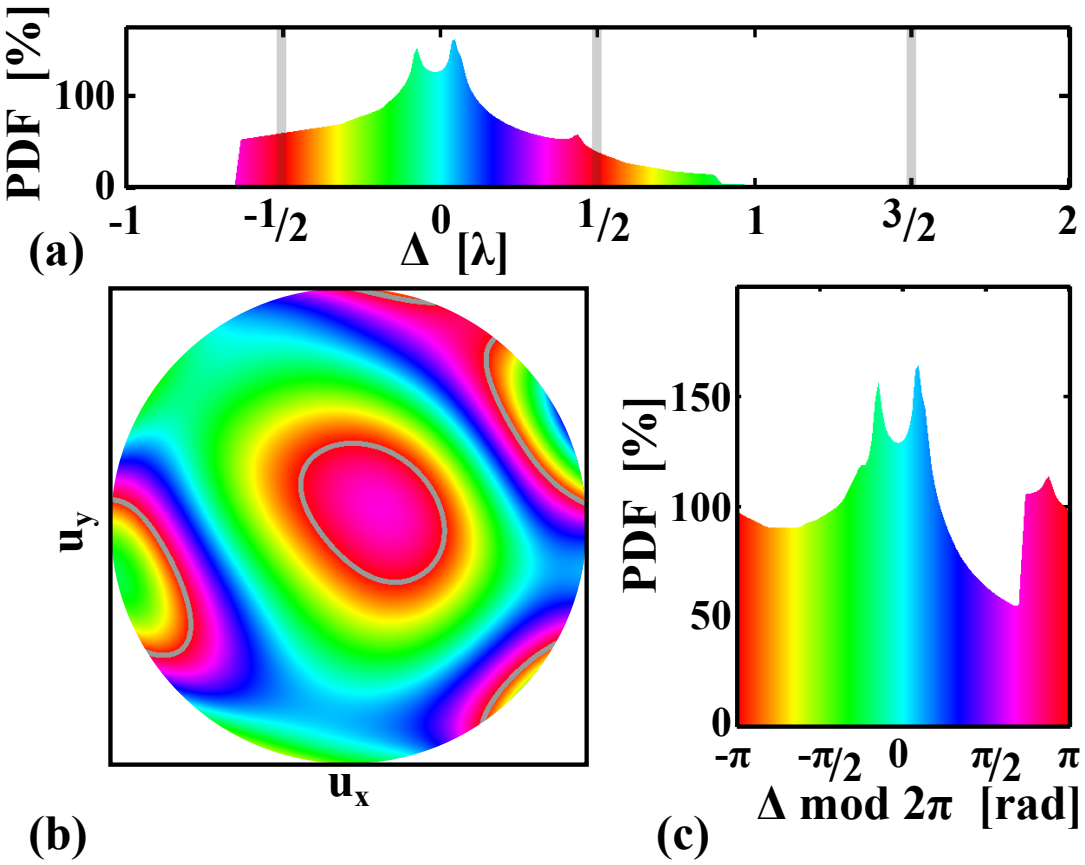


Figure 5.8: *The optical-path difference, $l(\mathbf{u})$, for a modest aberration with peak-to-valley optical-path difference of $\approx 1.5\lambda$, depicted as a probability density function (a), of its value in the pupil (b), and as a probability density function of its value $\varphi(\mathbf{u}) = l(\mathbf{u}) \mod 2\pi$. Lines of equal phase are indicated in grey, and the hue indicates the phase, $\varphi(\mathbf{u})$.*

As can be seen from figure 5.9c, the PDF of $\varphi(\mathbf{u})$ approximates a uniform distribution. Its convergence can be demonstrated mathematically following a similar derivation as that used in section 5.4.1 to show the uniformity of the PDF of $\Delta(\mathbf{u})$.

Furthermore, $\varphi_1(\mathbf{u})$ and $\varphi_2(\mathbf{u})$ will be independent when the phase difference $l(\mathbf{u}+\boldsymbol{\nu}) - l(\mathbf{u}-\boldsymbol{\nu}) \mod 2\pi$ is uniformly distributed. Hence, for sufficiently large aberrations, equation (5.2) can be approximated by:

$$OTF(\boldsymbol{\nu}) \approx \frac{1}{\operatorname{supp}(P)} \int_{-\pi}^{\pi} \frac{\mathrm{d}\varphi_1}{2\pi} \iint_{\Omega(\boldsymbol{\nu})} \exp(-\mathrm{i}\,\varphi_1) \exp(\mathrm{i}\, l(\mathbf{u}+\boldsymbol{\nu}))\, d\mathbf{u}, \tag{5.13}$$

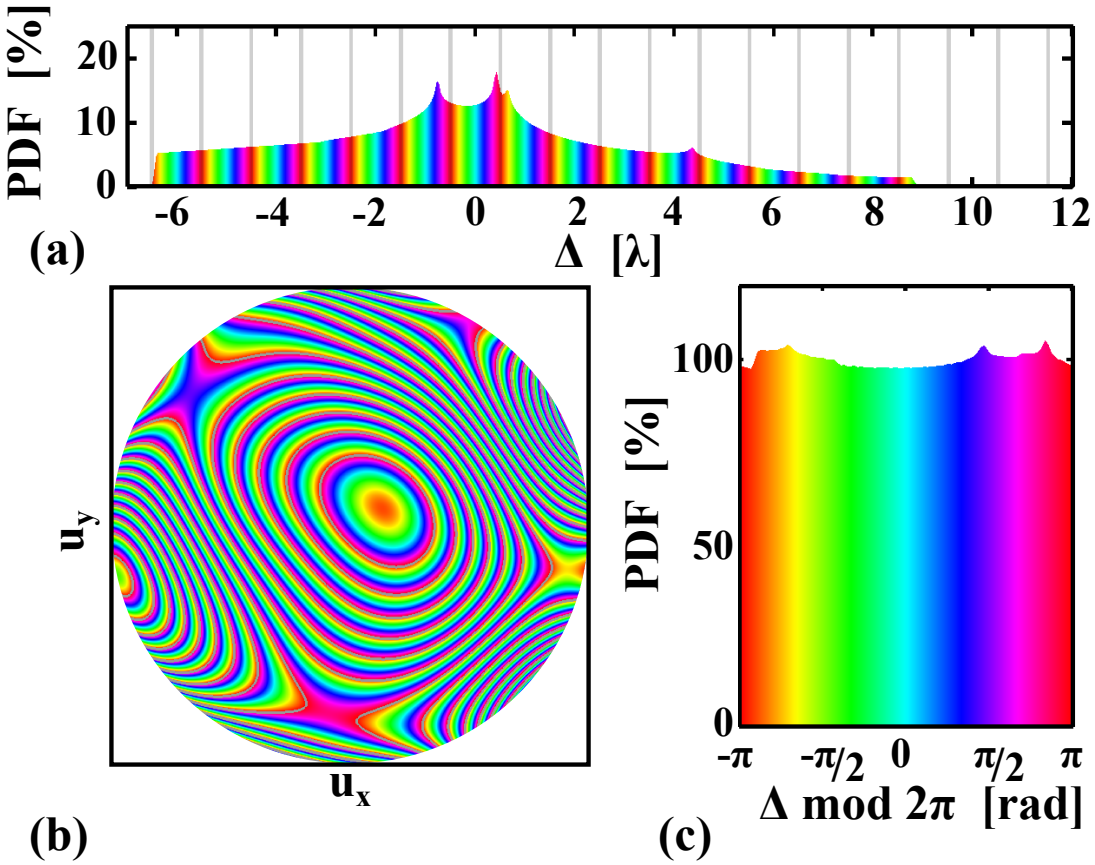


Figure 5.9: *The optical-path difference, $l(\mathbf{u})$, for a ten times larger aberration with peak-to-valley optical-path difference of $\approx 15\lambda$, depicted as a probability density function (a), of its value in the pupil (b), and as a probability density function of its value $\varphi(\mathbf{u}) = l(\mathbf{u}) \mod 2\pi$. Lines of equal phase are indicated in grey, and the hue indicates the phase, $\varphi(\mathbf{u})$. The optical-path difference, $l(\mathbf{u})$ of a large aberration can be seen to approximate a uniform distribution.*

permitting the separation of the first factor out of the internal double integral:

$$OTF(\boldsymbol{\nu}) \approx \frac{1}{\operatorname{supp}(P)} \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(-\mathrm{i}\,\varphi_1) \iint_{\Omega(\boldsymbol{\nu})} \exp(\mathrm{i}\,l(\mathbf{u}+\boldsymbol{\nu}))\, d\mathbf{u} \mathrm{d}\varphi_1 \tag{5.14}$$

Following the same procedure for $l(\mathbf{u}+\boldsymbol{\nu})$ as used earlier for $\Delta(\mathbf{u},\boldsymbol{\nu})$, the internal double integral can be converted to a single integration:

$$OTF(\boldsymbol{\nu}) \approx \frac{1}{\operatorname{supp}(P)} \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(-\mathrm{i}\,\varphi_1) \frac{A_\Omega(\boldsymbol{\nu})}{2\pi} \int_{-\pi}^{+\pi} \exp(\mathrm{i}\,\varphi_2)\, \mathrm{d}\varphi_2 \mathrm{d}\varphi_1 \tag{5.15}$$

which is equivalent to equation (5.6).

5.5 Conclusions

Binary masks can be designed to correct for phase aberrations to yield, in general, a contrast of approximately 20% ($2/\pi^2$) of the diffraction-limited MTF, and for a single spatial-frequency the relative contrast can in principle be as high as 64% ($2/\pi$). The absence of nulls and the relatively modest reduction in the MTF allows for digital recovery of a high quality image. Although the benefits of the masks are most pronounced for monochromatic imaging, good performance with an extended bandwidth is shown to be possible. By employing low dimensional mask-optimisation (in contrast to the high-dimensional optimisation proposed in [144]), binary contour masks can be calculated efficiently facilitating their use in adaptive optics. Furthermore, adaptive programmable masks could enable image recovery for unknown aberrations following a similar approach to those currently used in conjunction with deformable mirrors [147–149].

Chapter 6 – Conclusions and outlook

This thesis investigates how hybrid optical-digital imaging systems can be designed to yield the maximum image-fidelity in the presence of aberrations. In contrast to traditionally designed systems, hybrid imaging systems do not require a sharp intermediate optical image as long as sufficient information is captured to enable accurate image reconstruction. The extra degree of design freedom can be exploited to increase the tolerance to aberrations beyond what can be achieved with diffraction-limited optics and digital image processing.

Several types of pupil modulation have been reported to enable the aberration-tolerant transfer of sufficient information for the digital reconstruction of an image [18, 64, 19, 50, 54, 115, 53, 57, 60, 89, 52]. At the outset of this research it was unclear how different types of pupil-modulation compare. The problem is aggravated by the fact that it was unclear how large the pupil-modulation should be to provide the optimal image-fidelity. A sufficiently large modulation magnitude is required for acceptable aberration-tolerance; however, an excessively large modulation magnitude leads to a poor signal-to-noise ratio and image-fidelity.

To alleviate this situation, a method is proposed in this thesis to predict the imaging fidelity of hybrid imaging systems in the presence of aberrations. The method can be applied to compare arbitrary forms of pupil-modulation, and is used to evaluate common types of phase-modulation for their ability to extend the depth-of-focus of a hybrid imaging system. For this analysis; with the exception of defocus, the imaging system is free of aberrations, and able to restore the photon-shot-noise-limited image with the optimal mean-square minimising digital filter, i.e. the kernel used in image restoration is optimally matched to the actual PSF for each level of aberration.

It has been shown in this thesis that two local optimal generalised-cubic phase modulations minimise the expected imaging error of a defocus-tolerant hybrid imaging system; the global optimum closely resembles the original cubic phase modulation proposed by Dowski and Cathey [18], and the secondary optimum closely resembles the trefoil aberration. A modestly larger, albeit acceptable, imaging error is found for the radial-symmetric quartic phase modulation [57]. A second important observation made in this thesis is that the magnitude of the three optimal phase modulations increases approximately proportional with the maximum amount of defocus that is to be tolerated, and that this modulation magnitude is considerably lower than the values previously reported in the literature. It is shown in this thesis that in order to obtain the lowest expected imaging error, some variation of the image-fidelity with defocus must be accepted to avoid excessive suppression of the MTF. It has been shown in this thesis that a careful selection of the modulation magnitude to match the required defocus-tolerance can significantly reduce the expected imaging error.

The proposed method to evaluate hybrid imaging systems can be readily applied to more realistic optical and sensor models. In chapter 3 an efficient algorithm for the ray-trace-based calculation of the PSF is proposed, which facilitates the estimation of the expected imaging error iteratively as part of an optimisation or tolerancing process. It has also been shown in this thesis that the effects of sampling can be accounted for so that the imaging error can be estimated for hybrid imaging system that exhibit aliasing. Although the expected imaging error is larger for an under-sampled defocus-tolerant imaging system, the optimal magnitude of the cubic phase-modulation is seen to increase only by a modest amount. It has been noted in section 3.5.3 that in an otherwise axially-symmetric lens system, the orientation of a cubic phase-modulation with respect to the detector array has an influence on the expected imaging error.

It has been noted that the accurate simulation of broadband illumination could require a more dense sampling of the spectrum for pupil-engineered optics. Using an approximation of its OTF, an expression has been derived for the required number of wavelengths of the cubic phase modulation, which showed that for this example nearly 40 wavelengths are required to guarantee an accurate calculation of the OTF at the highest spatial frequencies. However, due to under-sampling and the dominance of the lower spatial frequencies in the power spectrum of the scene, the highest frequencies are typically of less importance to the image fidelity. When only the lower half or quarter of the spatial-frequency spectrum is considered, a more practical six or two wavelengths are respectively found to be sufficient.

The practicality of the proposed method has been demonstrated by integrating general-purpose computing software with commercial off-the-shelf optical-design software. Six of the most common types of phase modulation are optimised to yield the lowest hybrid imaging error when introduced in a realistic design of an uncooled thermal imaging system with an extended depth-of-field. A conservative level of noise, a realistic pixel fill-factor, and considerable aliasing as common in such imaging systems, have been accounted for. Instead of using the optimal mean-square error minimising filter, the image restoration is done using a simple Wiener deconvolution calculated for the in-focus PSF.

The optimal magnitude of most phase modulations was found to be comparable to that when using the optimal mean-square minimising digital filter. As we already predicted for the ideal system, the cubic phase modulation is also found to be optimal for the under-sampled system using the simple Wiener deconvolution. However, it was seen that the radially-symmetric phase modulations yield a lower imaging error than the trefoil modulation when employing a simple Wiener deconvolution.

The logarithmic axicon and the logarithmic asphere are seen to converge to the quartic phase modulation which in Chapter 4 have been shown to yield the lowest expected imaging error of the radially-symmetric phase-modulations. The optimisation of the

quartic phase modulation is also more efficient since its associated cost function is less-irregular than that of the phase functions containing a logarithm. Instead of evaluating several radially-symmetric modulation types, it is thus more practical to optimise the quartic phase modulation directly, perhaps amended with higher order terms.

Equipping commercial optical-design software with a hybrid imaging metric can enable the design of novel imaging systems. It has been shown in this thesis how a compact wide-field reflective system was optimised to mitigate off-axis aberrations and defocus across the full field-of-view. Acceptable image quality was only obtained by optimisation of various lens parameters in tandem with the generalised-cubic modulation of the primary mirror surface. This emphasises the importance of a hybrid imaging metric is central in the design process.

Although the highest light-throughput can be achieved with pupil-phase modulation, its implementation is typically more expensive than that of amplitude modulation, in particular when dynamic modulation is required. In contrast, efficient binary amplitude modulation can often be achieved at a lower cost. It has been shown here that binary masks can be optimised to correct for phase aberrations. The absence of nulls and the relatively modest reduction of the MTF permit the digital recovery of a high quality image. An expression has been derived for the upper limit of the MTF that can be obtained when masking an arbitrarily large aberration, as well as an expression for the MTF that is obtained when using the proposed *contour* mask. Although the benefits of contour masks are most pronounced for monochromatic imaging, good performance with an extended bandwidth has been shown to be possible with the optimisation of the hybrid imaging fidelity. By employing low dimensional mask-optimisation, binary contour masks can be calculated efficiently facilitating their use in adaptive optics.

Different types of pupil-modulation can be employed to improve the aberration-tolerance of a hybrid optical-digital imaging system. The method we proposed for the analysis of such systems allows finding the optimal parameters for many types of pupil modulation. In principle, the same technique can be used to find the optimal type of pupil function, even if no analytical description is available. This can be achieved by discretising the pupil function in a large number of small areas that are modulated by independent variables. Although with current computing hardware it is unlikely that global optimisation will yield the global optimum in an acceptable time, a high-dimensional local optimisation could be feasible. Such a local optimisation could be initiated from the optimum of one of the pupil-modulation types discussed in this thesis.

References

[1] N. J. Wade and S. Finger. The eye as an optical instrument: from camera obscura to helmholtz's perspective. *Perception*, 30(10):1157–1177, 2001.

[2] P. Noble. Self-scanned silicon image detector arrays. *Electron Devices, IEEE Transactions on*, 15(4):202–209, Apr 1968. ISSN 0018-9383.

[3] W. S. Boyle and G. E. Smith. Charge coupled semiconductor devices. *Bell Sys. Tech. J.*, 49(4):587–593, 1970.

[4] G. F. Amelio, M. F. Tompsett, and G. E. Smith. Experimental verification of the charge coupled device concept. *Bell Sys. Tech. J.*, 49(4):593–600, 1970. ISSN 0005-8580.

[5] J. N. Robert. Zernike polynomials and atmospheric turbulence. *JOSA*, 66(3): 207–211, Mar 1976. doi: 10.1364/JOSA.66.000207.

[6] D. Fried. Probability of getting a lucky short-exposure image through turbulence. *JOSA*, 68(12):1651–1657, 1978.

[7] C. Paterson, I. Munro, and J. Dainty. A low cost adaptive optics system using a membrane mirror. *Opt. Express*, 6(9):175–185, 2000.

[8] J. Liang and D. R. Williams. Aberrations and retinal image quality of the normal human eye. *JOSA A*, 14(11):2873–2883, 1997.

[9] S. Marcos, R. Navarro, and P. Artal. Coherent imaging of the cone mosaic in the living human eye. *JOSA A*, 13(5):897–905, 1996.

[10] J. Porter, A. Guirao, I. G. Cox, and D. R. Williams. Monochromatic aberrations of the human eye in a large population. *JOSA A*, 18(8):1793–1803, 2001.

[11] M. T. Sheehan, A. V. Goncharov, V. M. O'Dwyer, V. Toal, and C. Dainty. Population study of the variation in monochromatic aberrations of the normal human eye over the central visual field. *Opt. Express*, 15(12):7367–7380, 2007.

[12] C. Torti, S. Gruppetta, and L. Diaz-santana. Wavefront curvature sensing for the human eye. *J. Mod. Optic.*, 55(4):691–702, 2008. ISSN 0950-0340.

[13] M. R. Whalen. Correcting variable third-order astigmatism introduced by conformal aspheric surfaces. In *International Optical Design Conference*, volume 3482 of *Proc. SPIE*, pages 62–73, Sept. 1998.

[14] S. W. Sparrold, D. J. Knapp, P. K. Manhart, and K. W. Elsberry. Capabilities of an arch element for correcting conformal optical domes. In *Current developments in optical design and optical engineering VIII*, volume 3779, pages 434–444, Denver, Colorado, Jul 1999. Proc. SPIE.

[15] S. W. Sparrold, J. P. Mills, D. J. Knapp, K. S. Ellis, T. A. Mitchell, and P. K. Manhart. Conformal dome correction with counterrotating phase plates. *Optical Engineering*, 39:1822–1829, July 2000.

[16] G. E. Moore. Cramming more components onto integrated circuits. *Electronics Magazine*, 38(8):4, Apr 1965.

[17] G. Häusler. A method to increase the depth of focus by two step image processing. *Opt. Commun.*, 6(1):38–42, 1972.

[18] E. R. Dowski and W. T. Cathey. Extended depth of field through wave-front coding. *Appl. Opt.*, 34(11):1859–1866, 1995.

[19] S. Prasad, V. P. Pauca, R. J. Plemmons, T. C. Torgersen, and J. van der Gracht. Pupil-phase optimization for extended-focus, aberration-corrected imaging systems. In *Advanced Signal Processing Algorithms, Architectures, and Implementations XIV*, volume 5559 of *Proc. SPIE*, pages 335–345, 2004. doi: 10.1117/12.560235.

[20] A. Ashok and M. A. Neifeld. Point spread function engineering for iris recognition system design. *Appl. Opt.*, 49(10):B26–B39, Apr 2010. doi: 10.1364/AO.49.000B26.

[21] P. A. Jansson. *Deconvolution of images and spectra (2nd ed.)*. Academic Press, Inc., Orlando, FL, USA, 1996. ISBN 0-12-380222-9.

[22] J. van der Gracht, V. P. Pauca, H. Setty, R. Narayanswamy, R. J. Plemmons, S. Prasad, and T. Torgersen. Iris recognition with enhanced depth-of-field image acquisition. In *Visual Information Processing XIII*, volume 5438 of *Proc. SPIE*, pages 120–129, 2004.

[23] K. S. Kubala, H. B. Wach, V. V. Chumachenko, and E. R. Dowski. Increasing the depth of field in an lwir system for improved object identification. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVI*, volume 5784 of *Proc. SPIE*, pages 146–156, May 2005. doi: 10.1117/12.604079.

[24] R. Narayanswamy and P. E. X. Silveira. Iris recognition at a distance with expanded imaging volume. In *Biometric Technology for Human Identification III*, volume 6202 of *Proc. SPIE*, page 62020G, 2006.

[25] E. H. Adelson and J. Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:99–106, 1992. ISSN 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/34.121783.

[26] R. N. M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. Computer Science Tech Report CSTR 02, Stanford University, April 2005.

[27] R. Ng. *Digital light field photography*. PhD thesis, School of Engineering, Stanford, CA, USA, 2006. AAI3219345.

[28] W. T. Welford. Use of annular apertures to increase focal depth. *JOSA*, 50(8): 749–752, 1960.

[29] M. Mino and Y. Okano. Improvement in the otf of a defocused optical system through the use of shaded apertures. *Appl. Opt.*, 10(10):2219–2225, October 1971.

[30] J. Ojeda-Castaneda, R. Ramos, and A. Noyola-Isgleas. High focal depth by apodization and digital restoration. *Appl. Opt.*, 27(12):2583–2586, June 1988.

[31] J. Ojeda-Castaneda, E. Tepichin, and A. Pons. Apodization of annular apertures: Strehl ratio. *Appl. Opt.*, 27(24):5140+, December 1988.

[32] J. D. Kraus. *Radio Astronomy*. Cygnus–Quasar Books, Powell, Ohio, 1986.

[33] J. E. Baldwin, C. A. Haniff, C. D. Mackay, and P. J. Warner. Closure phase in high-resolution optical imaging. *Nature*, 320:595–597, Apr. 1986. doi: 10.1038/320595a0.

[34] T. J. Cornwell. The applications of closure phase to astronomical imaging. *Science*, 245(4915):263–269, Jul 21 1989. ISSN 0036-8075.

[35] J. J. Alleman, F. Reynaud, and P. Connes. Fiber-linked telescope array: description and laboratory tests of a two-channel prototype. *Appl. Opt.*, 34(13):2284, 1995.

[36] C. A. Haniff, C. D. Mackay, D. J. Titterington, D. Sivia, and J. E. Baldwin. The first images from optical aperture synthesis. *Nature*, 328:694–696, Aug. 1987. doi: 10.1038/328694a0.

[37] J. S. Young, J. E. Baldwin, R. C. Boysen, C. A. Haniff, P. R. Lawson, C. D. Mackay, D. Pearson, J. Rogers, D. St.-Jacques, P. J. Warner, D. M. A. Wilson, and R. W. Wilson. New views of Betelgeuse: multi-wavelength surface imaging and implications for models of hotspot generation. *Mon. Not. R. Astron. Soc.*, 315:635–645, July 2000. doi: 10.1046/j.1365-8711.2000.03438.x.

[38] P. G. Tuthill, J. D. Monnier, W. C. Danchi, E. H. Wishnow, and C. A. Haniff. Michelson interferometry with the keck i telescope. *Publ. Astron. Soc. Pac.*, 112: 555–565, Apr. 2000. doi: 10.1086/316550.

[39] J. P. Lloyd, F. Martinache, M. J. Ireland, J. D. Monnier, S. H. Pravdo, S. B. Shaklan, and P. G. Tuthill. Direct detection of the brown dwarf GJ 802B with adaptive optics masking interferometry. *Astrophys. J.*, 650(2, Part 2):L131–L134, OCT 20 2006. ISSN 0004-637X.

[40] C. W. Slinger, K. Gilholm, N. Gordon, M. McNie, D. Payne, K. Ridley, M. Strens, M. Todd, G. De Villiers, P. Watson, R. Wilson, G. Dyer, M. Eismann, J. Meola, and S. Rogers. Adaptive coded aperture imaging in the infrared: towards a practical implementation. In *Adaptive Coded Aperture Imaging and Non-Imaging Sensors II*, volume 7096 of *Proc. SPIE*, Aug. 2008. doi: 10.1117/12.805372.

[41] J. G. Ables. Fourier transform photography: A new method for x-ray astronomy. In *Proc. Astron. Soc. Aust. 4, 172-3*, 1968.

[42] R. H. Dicke. Scatter-hole cameras for x-rays and gamma rays. *Astrophys. J.*, 153: 101–106, 1968.

[43] C. Slinger, N. Gordon, K. Lewis, G. McDonald, M. McNie, D. Payne, K. Ridley, M. Strens, G. De Villiers, and R. Wilson. Coded aperture systems as non-conventional lensless imagers for the visible and infrared. In *Electro-Optical and Infrared Systems: Technology and Applications IV.*, volume 6737 of *Proc. SPIE*, page 67370D, Nov. 2007. doi: 10.1117/12.739043.

[44] C. Slinger, M. Eismann, N. Gordon, K. Lewis, G. McDonald, M. McNie, D. Payne, K. Ridley, M. Strens, G. De Villiers, and R. Wilson. An investigation of the potential for the use of a high resolution adaptive coded aperture system in the mid-wave infrared. In *Adaptive Coded Aperture Imaging and Non-Imaging Sensors*, volume 6714 of *Proc. SPIE*, Oct. 2007. doi: 10.1117/12.736071.

[45] J. W. Goodman. *Introduction to Fourier Optics*. Roberts & Company Publishers, December 2005. ISBN 0974707724.

[46] R. P. Millane and W. H. Hsiao. The basis of phase dominance. *Opt. Lett.*, 34(17): 2607–2609, Sep 2009. doi: 10.1364/OL.34.002607.

[47] M. Demenikov and A. R. Harvey. Image artifacts in hybrid imaging systems with a cubic phase mask. *Opt. Express*, 18(8):8207–8212, 2010.

[48] Z. Zalevsky, A. Shemer, A. Zlotnik, E. B. Eliezer, and E. Marom. All-optical axial super resolving imaging using a low-frequency binary-phase mask. *Opt. Express*, 14(7):2631–2643, 2006.

[49] A. Zlotnik, S. B. Yaish, O. Yehezkel, K. Lahav-Yacouel, M. Belkin, and Z. Zalevsky. Extended depth of focus contact lenses for presbyopia. *Opt. Lett.*, 34(14):2219–2221, Jul 2009. doi: 10.1364/OL.34.002219.

[50] W. Chi and N. George. Electronic imaging using a logarithmic asphere. *Opt. Lett.*, 26(12):875–877, 2001.

[51] J. McLeod. Axicons and their uses. *JOSA*, 50(2):166–166, 1960.

[52] S. Mezouari and A. R. Harvey. Phase pupil functions for reduction of defocus and spherical aberrations. *Opt. Lett.*, 28(10):771–773, 2003.

[53] J. Sochacki, S. Bara, Z. Jaroszewicz, and A. Kolodziejczyk. Phase retardation of the uniform-intensity axilens. *Opt. Lett.*, 17(1):7, 1992.

[54] W. Chi and N. George. Computational imaging with the logarithmic asphere: theory. *JOSA A*, 20(12):2260–2273, 2003. doi: 10.1364/JOSAA.20.002260.

[55] A. Sauceda and J. Ojeda-Castañeda. High focal depth with fractional-power wave fronts. *Opt. Lett.*, 29(6):560–562, 2004.

[56] D. Zalvidea and E. E. Sicre. Phase pupil functions for focal-depth enhancement derived from a wigner distribution function. *Appl. Opt.*, 37(17):3623–3627, 1998.

[57] D. Zalvidea, C. Colautti, and E. E. Sicre. Quality parameters analysis of optical imaging systems with enhanced focal depth using the wigner distribution function. *JOSA A*, 17(5):867–873, 2000.

[58] M. D. Robinson, G. Feng, and D. G. Stork. Spherical coded imagers: improving lens speed, depth-of-field, and manufacturing yield through enhanced spherical aberration and compensating image processing. In *Novel Optical Systems Design and Optimization XII*, volume 7429 of *Proc. SPIE*, pages 74290–M, 2009.

[59] D. Robinson and D. Stork. Extending depth-of-field: Spherical coding versus asymmetric wavefront coding. In *Computational Optical Sensing and Imaging (COSI)*, page CThB3. OSA Technical Digest (CD), 2009.

[60] S. Mezouari, G. Muyo, and A. R. Harvey. Circularly symmetric phase filters for control of primary third-order aberrations: coma and astigmatism. *JOSA A*, 23 (5):1058–1062, 2006.

[61] Q. Yang, L. Liu, and J. Sun. Optimized phase pupil masks for extended depth of field. *Opt. Commun.*, 272(1):56 – 66, 2007. ISSN 0030-4018. doi: 10.1016/j.optcom.2006.11.021.

[62] N. Caron and Y. Sheng. Polynomial phase masks for extending the depth of field of a microscope. *Appl. Opt.*, 47(22):E39–E43, 2008.

[63] S. S. Sherif, T. W. Cathey, and E. R. Dowski. Phase plate to extend the depth of field of incoherent hybrid imaging systems. *Appl. Opt.*, 43(13):2709–2721, 2004.

[64] S. Prasad, T. C. Torgersen, V. P. Pauca, R. J. Plemmons, and J. van der Gracht. Engineering the pupil phase to improve image quality. In *Visual Information Processing XII*, volume 5108 of *Proc. SPIE*, pages 1–12, 2003. doi: 10.1117/12.487572.

[65] Y. Frauel and A. Castro. Increasing the depth of field of imaging systems with numerically optimized phase masks. In *Security and Defence 2007 (Florence, Italy)*, volume 6736 of *Proc. SPIE*, 2007. doi: 10.1117/12.740506.

[66] A. Ashok and M. A. Neifeld. Pseudorandom phase masks for superresolution imaging from subpixel shifting. *Appl. Opt.*, 46(12):2256–2268, 2007.

[67] Q. Yang, L. Liu, and H. Lang. Enlarging the depth of focus by filtering in the phase-space domain. *Appl. Opt.*, 44(32):6833–6840, Nov 2005. doi: 10.1364/AO.44.006833.

[68] S. Mezouari, G. Muyo, and A. Harvey. Amplitude and phase filters for mitigation of defocus and third-order aberrations. In *Optical Design and Engineering*, volume 5249 of *Proc. SPIE*, pages 238–248, 2004. ISBN 0-8194-5133-9.

[69] E. Ben-Eliezer, N. Konforti, B. Milgrom, and E. Marom. An optimal binary amplitude-phase mask for hybrid imaging systems that exhibit high resolution and extended depth of field. *Opt. Express*, 16(25):20540–20561, 2008.

[70] S. Bagheri and B. Javidi. Extension of depth of field using amplitude and phase modulation of the pupil function. *Opt. Lett.*, 33(7):757–759, 2008.

[71] S. Abrahamsson, S. Usawa, and M. Gustafsson. A new approach to extended focus for high-speed, high-resolution biological microscopy. In *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XIII*, volume 6090 of *Proc. SPIE*, page 60900N, 2006. doi: 10.1117/12.647022.

[72] W. Chi, K. Chu, and N. George. Polarization coded aperture. *Opt. Express*, 14 (15):6634–6642, 2006.

[73] E. H. Linfoot and E. Wolf. Diffraction images in systems with an annular aperture. *Proceedings of the Physical Society. Section B*, 66(2):145–149, 1953. doi: 10.1088/0370-1301/66/2/312.

[74] M. Born and E. Wolf. *Principles of Optics, Electromagnetic Theory of Propagation, Interference and Diffraction of Light (Sixth Edition)*. Pergamon Press, 1980.

[75] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[76] K. Lee, J. Chang, and J. Yoon. High performance microshutter device with space-division modulation. *J. Micromech. Microeng.*, 20(7):075030, 2010.

[77] S. Goodwin, B. Stoner, J. Carlson, and S. Rogers. Artificial eyelid dynamic aperture optical arrays for large scale coding elements with application in visible to mwir. In *Adaptive Coded Aperture Imaging and Non-Imaging Sensors II*, volume 7096 of *Proc. SPIE*, pages 70960E–10, Bellingham, WA, USA, 2008.

[78] A. Kutyrev, R. Arendt, S. Moseley, R. Boucarut, T. Hadjimichael, M. Jhabvala, T. King, M. Li, J. Loughlin, D. Rapchun, D. Schwinger, and R. Silverberg. Programmable microshutter arrays for the JWST NIRSpec: optical performance. *IEEE J. Sel. Top. Quant.*, 10(3):652–661, 2004.

[79] M. McNie, D. King, N. Price, D. Combes, G. Smith, A. Brown, N. Gordon, S. Stone, K. Brunson, K. Lewis, C. Slinger, and S. Rogers. A large area reconfigurable moems microshutter array for coded aperture imaging systems. In *Adaptive Coded Aperture Imaging and Non-Imaging Sensors II*, volume 7096 of *Proc. SPIE*, pages 70960D–8, 2008.

[80] E. R. Dowski and G. E. Johnson. Wavefront coding: a modern method of achieving high-performance and/or low-cost imaging systems. In *Current Developments in Optical Design and Optical Engineering VIII*, volume 3779 of *Proc. SPIE*, pages 137–145, Oct. 1999.

[81] R. Narayanswamy, G. Johnson, P. X. Silveira, and H. Wach. Extending the imaging volume for biometric iris recognition. *Appl. Opt.*, 44(5):701–712, Feb 2005. doi: 10.1364/AO.44.000701.

[82] P. E. X. Silveira and R. Narayanswamy. Signal-to-noise analysis of task-based imaging systems with defocus. *Appl. Opt.*, 45(13):2924–2934, May 2006. doi: 10.1364/AO.45.002924.

[83] B. Milgrom, N. Konforti, M. A. Golub, and E. Marom. Novel approach for extending the depth of field of barcode decoders by using RGB channels of information. *Opt. Express*, 18(16):17027–17039, Aug 2010. doi: 10.1364/OE.18.017027.

[84] P. J. Bex and W. Makous. Spatial frequency, phase, and the contrast of natural images. *JOSA A*, 19(6):1096–1106, 2002.

[85] E. P. Simoncelli. *Statistical Modeling of Photographic Images*, chapter 4.7, pages 431–441. Academic Press, May 2005. 2nd edition.

[86] V. A. Billock. Neural acclimation to 1/f spatial frequency spectra in natural images transduced by the human visual system. *Phys. D*, 137(3-4):379–391, 2000. ISSN 0167-2789. doi: 10.1016/S0167-2789(99)00197-9.

[87] D. L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.*, 73(6):814–817, Aug 1994. doi: 10.1103/PhysRevLett.73.814.

[88] A. van der Schaaf and J. H. van Hateren. Modelling the power spectra of natural images: Statistics and information. *Vision Res.*, 36(17):2759–2770, 1996. ISSN 0042-6989.

[89] S. Mezouari and A. R. Harvey. Combined amplitude and phase filters for increased tolerance to spherical aberration. *J. Mod. Optic.*, 50:2213–2220, Jan. 2003. doi: 10.1080/0950034031000099205.

[90] P. Mouroulis. Depth of field extension with spherical optics. *Opt. Express*, 16 (17):12995–13004, Aug 2008. doi: 10.1364/OE.16.012995.

[91] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari. Tikhonov regularization and the l-curve for large discrete ill-posed problems. *J. Comput. Appl. Math.*, 123 (1-2):423–446, 2000. ISSN 0377-0427. doi: 10.1016/S0377-0427(00)00414-3.

[92] W. H. Richardson. Bayesian-based iterative method of image restoration. *JOSA*, 62(1):55–59, 1972.

[93] L. B. Lucy. An iterative technique for the rectification of observed distributions. *AJ*, 79(6):745–754, June 1974.

[94] G. Muyo, A. Singh, M. Andersson, D. Huckridge, A. Wood, and A. R. Harvey. Infrared imaging with a wavefront-coded singlet lens. *Opt. Express*, 17(23): 21118–21123, 2009.

[95] M. Demenikov, E. Findlay, and A. R. Harvey. Miniaturization of zoom lenses with a single moving element. *Opt. Express*, 17(8):6118–6127, 2009.

[96] J. Markham and J.-A. Conchello. Parametric blind deconvolution: a robust method for the simultaneous estimation of image and blur. *JOSA A*, 16(10):2377–2391, 1999.

[97] Y.-L. You and M. Kaveh. Blind image restoration by anisotropic regularization. *Image Processing, IEEE Transactions on*, 8(3):396–407, Mar 1999. ISSN 1057-7149. doi: 10.1109/83.748894.

[98] M. Demenikov and A. R. Harvey. Parametric blind-deconvolution algorithm to remove image artifacts in hybrid imaging systems. *Opt. Express*, 18(17):18035–18040, 2010.

[99] G. Muyo and A. R. Harvey. Decomposition of the optical transfer function: wavefront coding imaging systems. *Opt. Lett.*, 30(20):2715–2717, 2005.

[100] J. Mannos and D. Sakrison. The effects of a visual fidelity criterion of the encoding of images. *Information Theory, IEEE Transactions on*, 20(4):525–536, Jul 1974. ISSN 0018-9448.

[101] C. J. van den Branden Lambrecht. A working spatio-temporal model of the human visual system for image restoration and quality assessment applications. In *Acoustics, Speech, and Signal Processing - IEEE International Conference, Volume 04*, ICASSP '96, pages 2291–2294, Washington, DC, USA, 1996. IEEE Computer Society. ISBN 0-7803-3192-3.

[102] G. Muyo and A. R. Harvey. The effect of detector sampling in wavefront-coded imaging systems. *J. Opt. A: Pure Appl. Opt.*, 11(5):054002 (9pp), MAY 2009. ISSN 1464-4258. doi: 10.1088/1464-4258/11/5/054002.

[103] E. M. Stein. *Harmonic Analysis: Real-variable Methods, Orthogonality, and Oscillatory Integrals (with the assistance of Timothy S. Murphy)*. Princeton University Press, Princeton, NJ, 1993.

[104] D. Huybrechs and S. Vandewalle. The construction of cubature rules for multivariate highly oscillatory integrals. *Math. Comp.*, 76(260):1955–1980, 2007.

[105] B. Richards and E. Wolf. Electromagnetic diffraction in optical systems. ii. structure of the image field in an aplanatic system. *Royal Society of London Proceedings Series A*, 253:358–379, Dec. 1959.

[106] F. B. Hildebrand. *Introduction to Numerical Analysis*. Courier Dover Publications, 1987.

[107] R. Wodnicki, G. W. Roberts, and M. D. Levine. A foveated image sensor in standard CMOS technology. In *Custom Integrated Circuits Conference, Proc. IEEE*, pages 357–360, 1995.

[108] R. H. Vollmerhausen and R. G. Driggers. *Analysis of Sampled Imaging Systems*. SPIE Press, 2000. ISBN 0-8194-3489-2.

[109] Z. Zalevsky and D. Mendlovic. *Optical superresolution*. Springer-Verlag, 2004.

[110] M. V. Berry and N. L. Balazs. Non-spreading wave packets. *Am. J. Phys.*, 47(3): 264–267, 1979. ISSN 0002-9505.

[111] M. A. Bandres and J. C. Gutiérrez-Vega. Airy-gauss beams and their transformation by paraxial optical systems. *Opt. Express*, 15(25):16719–16728, Dec 2007. doi: 10.1364/OE.15.016719.

[112] G. A. Siviloglou, J. Broky, A. Dogariu, and D. N. Christodoulides. Observation of accelerating airy beams. *Phys. Rev. Lett.*, 99(21):213901, Nov 2007. doi: 10.1103/PhysRevLett.99.213901.

[113] P. Polynkin, M. Kolesik, J. V. Moloney, G. A. Siviloglou, and D. N. Christodoulides. Curved plasma channel generation using ultraintense airy beams. *Science*, 324(5924):229–232, 2009. doi: 10.1126/science.1169544.

[114] T. Vettenburg, N. Bustin, and A. R. Harvey. Fidelity optimization for aberration-tolerant hybrid imaging systems. *Opt. Express*, 18(9):9220–9228, 2010.

[115] N. Davidson, A. A. Friesem, and E. Hasman. Holographic axilens: high resolution and long focal depth. *Opt. Lett.*, 16(7):523–525, 1991.

[116] G. A. Siviloglou and D. N. Christodoulides. Accelerating finite energy airy beams. *Opt. Lett.*, 32(8):979–981, 2007.

[117] R. Storn and K. Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.*, 11(4):341–359, 1997. ISSN 0925-5001.

[118] J. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[119] J. Durnin, J. J. Miceli, and J. H. Eberly. Diffraction-free beams. *Phys. Rev. Lett.*, 58(15):1499–1501, Apr 1987. doi: 10.1103/PhysRevLett.58.1499.

[120] G. D. Muyo and A. R. Harvey. Wavefront coding for athermalization of infrared imaging systems. In *Electro-Optical and Infrared Systems: Technology and Applications*, volume 5612 of *Proc. SPIE*, pages 227–235, Oct. 2004. doi: 10.1117/12.579738.

[121] G. Muyo, A. Singh, M. Andersson, D. Huckridge, and A. R. Harvey. Optimized thermal imaging with a singlet and pupil plane encoding: experimental realization. In *Electro-Optical and Infrared Systems: Technology and Applications III.*, volume 6395 of *Proc. SPIE*, page 63950M, Oct. 2006. doi: 10.1117/12.689765.

[122] M. Demenikov, E. Findlay, and A. Harvey. Miniaturization and simplification of zoom lenses using wavefront coding. In *Novel Optical Systems Design and Optimization XI*, volume 7061A of *Proc. SPIE*, Aug. 2008.

[123] E. J. Tremblay, R. A. Stack, R. L. Morrison, and J. E. Ford. Ultrathin cameras using annular folded optics. *Appl. Opt.*, 46(4):463–471, 2007.

[124] E. J. Tremblay, R. A. Stack, R. L. Morrison, J. H. Karp, and J. E. Ford. Ultrathin four-reflection imager. *Appl. Opt.*, 48(2):343–354, 2009.

[125] F. Yan and X. Zhang. Optimization of an off-axis three-mirror anastigmatic system with wavefront coding technology based on mtf invariance. *Opt. Express*, 17(19):16809–16819, 2009.

[126] F. Yan and X.-j. Zhang. The effect on tolerance distributing of an off-axis three mirror anastigmatic optical system with wavefront coding technology. In *Optical System Alignment and Tolerancing II*, volume 7068 of *Proc. SPIE*, pages 706807–14, San Diego, CA, USA, 2008.

[127] A. S. Leonard. *The Yolo reflector*. OPTICA, 1965.

[128] A. Mackintosh, K. Hawkings, and F. Baar. *Optics*. Willmann-Bell, Inc., 1986.

[129] G. Kreider, J. V. der Spiegel, I. Born, C. L. Claeys, I. Debusschere, G. Sandini, and P. Dario. Design and characterization of a space-variant ccd sensor. In *Intelligent Robots and Computer Vision IX: Algorithms and Techniques*, volume 1381 of *Proc. SPIE*, pages 242–249, 1991.

[130] H. Trussell and B. Hunt. Sectioned methods for image restoration. *IEEE Trans. Acoust. Speech, Signal Process*, 26(2):157–164, 1978. doi: 10.1109/TASSP.1978.1163073.

[131] H. Trussell and B. Hunt. Image restoration of space-variant blurs by sectioned methods. *IEEE Trans. Acoust. Speech, Signal Process*, 26(6):608–609, 1978. doi: 10.1109/TASSP.1978.1163161.

[132] J. G. Nagy, V. P. Pauca, R. J. Plemmons, and T. C. Torgersen. Space-varying restoration of optical images. *JOSA A*, 14(12):3162–3174, 1997.

[133] J. G. Nagy and D. P. O'Leary. Fast iterative image restoration with a spatially varying PSF. In *Advanced Signal Processing: Algorithms, Architectures, and Implementations VII*, volume 3162 of *Proc. SPIE*, pages 388–399, San Diego, CA, USA, 1997.

[134] N. Hajlaoui, C. Chaux, G. Perrin, F. Falzon, and A. Benazza-Benyahia. Satellite image restoration in the context of a spatially varying point spread function. *JOSA A*, 27(6):1473–1481, Jun 2010. doi: 10.1364/JOSAA.27.001473.

[135] H. Hofer, L. Chen, G.-Y. Yoon, B. Singer, Y. Yamauchi, and D. R. Williams. Improvement in retinal image quality with dynamic correction of the eye's aberrations. *Opt. Express*, 8(11):631–643, 2001.

[136] E. J. Fernández, I. Iglesias, and P. Artal. Closed-loop adaptive optics in the human eye. *Opt. Lett.*, 26(10):746–748, 2001.

[137] C. Paterson, I. Munro, and J. Dainty. Low cost adaptive optics systems. In *Adaptive optical systems technology*, volume 4007 of *Proc. SPIE*, pages 185–193, 2000.

[138] S. E. Broomfield, M. A. A. Neil, E. G. S. Paige, and G. G. Yang. Programmable binary phase-only optical device based on ferroelectric liquid crystal SLM. *Electron. Lett.*, 28(1):26–28, 1992. ISSN 0013-5194.

[139] G. D. Love, N. Andrews, P. Birch, D. Buscher, P. Doel, C. Dunlop, J. Major, R. Myers, A. Purvis, R. Sharples, A. Vick, A. Zadrozny, S. R. Restaino, and A. Glindemann. Binary adaptive optics: atmospheric wave-front correction with a half-wave phase shifter. *Appl. Opt.*, 34(27):6058–6066, 1995.

[140] P. M. Birch, J. Gourlay, G. D. Love, and A. Purvis. Real-time optical aberration correction with a ferroelectric liquid-crystal spatial light modulator. *Appl. Opt.*, 37(11):2164–2169, 1998.

[141] G. D. Love. Binary adaptive optics: atmospheric wave-front correction with a half-wave phase shifter: addenda. *Appl. Opt.*, 35(3):347–347, 1996.

[142] J. Osborn, R. M. Myers, and G. D. Love. PSF halo reduction in adaptive optics using dynamic pupil masking. *Opt. Express*, 17(20):17279–17292, 2009.

[143] A. Kolmogorov and S. Fomin. *Introductory Real Analysis*. Dover, 1975.

[144] J. Stayman, N. Subotic, and W. Buller. An analysis of coded aperture acquisition and reconstruction using multi-frame code sequences for relaxed optical design constraints. In *Adaptive Coded Aperture Imaging, Non-Imaging, and Unconventional Imaging Sensor Systems*, volume 7468 of *Proc. SPIE*, pages 74680D–12, San Diego, CA, USA, 2009.

[145] A. Guirao, M. Redondo, and P. Artal. Optical aberrations of the human cornea as a function of age. *JOSA A*, 17(10):1697–1702, 2000.

[146] A. Gorman, D. Fletcher-Holmes, and A. Harvey. Generalization of the lyot filter and its application to snapshot spectral imaging. *Opt. Express*, 18(6):5602–5608, 2010.

[147] R. Muller and A. Buffington. Real-time correction of atmospherically degraded telescope images through image sharpening. *JOSA*, 64(9):1200–1210, 1974.

[148] A. Buffington, F. Crawford, R. Muller, A. Schwemin, and R. Smits. Correction of atmospheric distortion with an image-sharpening telescope. *JOSA*, 67(3):298–303, 1977.

[149] D. Débarre, M. Booth, and T. Wilson. Image based adaptive optics through optimisation of low spatial frequencies. *Opt. Express*, 15(13):8176–8190, 2007.